Check for updates

WILEY

Effects of speaker types and L1 backgrounds on the linguistic complexity of learners' writing

Hui Wang¹ | Gui Wang¹ | Nan Wang² | Li Wang³ \bigcirc

¹School of International Studies, Zhejiang University, Zhejiang, China

²School of Chinese Language and Literature, Beijing Foreign Studies University, Beijing, China

³Foreign Languages College, Shanghai Normal University, Shanghai, China

Correspondence

Li Wang, Foreign Languages College, Shanghai Normal University, 100 Guilin Road, Xuhui, Shanghai, China. Email: wanglily22@shnu.edu.cn

Funding information

National Social Science Foundation of China. Grant/Award Number: 17BYY115

Abstract

This study investigated the differences in linguistic complexity between different types of English users, including native speakers (NS), English as a foreign language (EFL) learners, and English as a second language (ESL) learners in terms of Kolmogorov complexity. Furthermore, we explored how first language backgrounds affect linguistic complexity. Our dataset contains 2272 argumentative essays produced by English NSs and upper-intermediate learners from four ESL and six EFL countries/regions. Results showed that significant differences existed between the writings of NS, EFL, and ESL regarding overall and syntactic complexity. Specifically, the rank of overall complexity (NS > ESL > EFL) indicates that learners from countries/regions with higher exposure to English tend to produce overall more complex writings. Concerning syntactic complexity, EFL learners produce the most complex writings, while NS produces the least complex, indicating that essays written by EFL learners contain the most fixed word order patterns. In contrast, no significant difference was detected in morphological complexity among the NS, ESL, and EFL groups, suggesting that native and upper-intermediate non-NSs exhibit a similar range of morphological forms in their writings. Additionally, our results showed a larger effect of first language backgrounds over English speaker types on linguistic complexity, thus informing teachers to implement targeted writing instructions for learners from different countries.

KEYWORDS

EFL learners, ESL learners, L1 backgrounds, learner corpus, linguistic complexity, second language writing

摘要

本研究旨在深入探讨英语母语者 (Native Speakers, NS)、英 语作为外语的学习者 (English as a Foreign Language learners, EFL learners) 以及英语作为第二语言的学习者 (English as a Second Language learners, ESL learners) 在柯尔莫哥洛 夫整体、句法及形态复杂度方面的差异性,并进一步剖析 学习者的母语背景对其语言复杂度的潜在影响。为此,本 文选取了2272篇来自英语母语者及四个ESL国家/地区和六 个EFL国家/地区的中高级学习者所撰写的议论文文章进行 深入分析。研究发现,在整体和句法复杂度上,三类英语使用 者之间存在显著性差异。具体表现为,整体复杂度方面,英语 母语者的输出超越了ESL和EFL学习者,暗示英语接触频率可 能与语言输出复杂度正相关。句法复杂度方面, EFL学习者 的输出呈现出最高的句法复杂度,而英语母语者则相对较低, 这可能与EFL学习者在写作中更为固定的词序模式有关。而 在形态复杂度上,三类英语使用者并未表现出显著差异,暗示 其在写作中所采用的词汇形态范围具有一致性。此外,相较 于英语使用者的类别,学习者的母语背景对其语言复杂度的 影响更为显著。鉴于此,教育者在为不同母语背景的学习者 提供写作指导时,应充分考虑其母语背景的潜在影响,以确保 教学策略的针对性和实效性。

关键词

英语作为外语学习者,英语作为二语学习者,母语背景,学习者语料 库,语言复杂度,二语写作

1 | INTRODUCTION

Recent years have witnessed a growing academic interest in linguistic complexity of second language (L2) writing research, as evidenced by a large number of relevant studies (Barrot & Gabinete, 2021; Biber et al., 2020; Ehret & Szmrecsanyi, 2019; Jiang et al., 2023; Khushik & Huhta, 2020; Kyle & Crossley, 2018). Some studies have proposed various complexity measures and investigated whether they are reliable in gauging learners' proficiency in the target language (Brezina & Pallotti, 2016; Kyle & Crossley, 2017, 2018; Lu, 2010, 2011; Ouyang et al., 2022). Others have explored certain tasks, contexts, and learner-related factors that influence linguistic complexity, such as genre (Yoon, 2021), planning time (Seyyedi et al., 2013), and learner's first language (L1) backgrounds (Khushik & Huhta, 2020).

Whereas these studies are certainly informative, most of them focus on syntactic complexity or lexical complexity, leaving morphological complexity rarely explored (Brezina & Pallotti, 2016; Martínez-Adrián, & Nieva-Marroquín, 2023; Yoon, 2018). Considering the multifaceted nature of L2 writing performance, such a gap may result in our limited understanding of variations of linguistic complexity in L2 writing.

Furthermore, little attention has been paid to the differences in linguistic complexity between different English speaker types, including native speakers (NS), English as a foreign language (EFL) learners, and English as a second language (ESL) learners. The existing literature investigating complexity variations among different types of English speakers either involved NS and merely one type of L2 learners (i.e., ESL/EFL) or considered all L2 learners from different L1 backgrounds as a whole group. As argued by Barrot and Gabinete (2021), NS, ESL, and EFL learners may possess their own distinct ways of producing the English language because of the different sociopolitical status and internal communicative functions of English. Thus, it would be essential to investigate the potential complexity differences among these three groups of English speakers.

To these ends, this study seeks to examine the differences in linguistic complexity between NS, EFL learners, and ESL learners by adopting Kolmogorov complexity, a holistic information-theoretic approach. Such a holistic metric can measure three facets of L2 writing performance simultaneously, namely, overall, syntactic, and morphological complexity. Taking into account these three sublevels of linguistic complexity, the present research will provide a more comprehensive perspective on complexity research in L2 writing performance, complementing the previous literature which focuses primarily on a certain sublevel.

Additionally, we went a further step to investigate how specific L1 backgrounds affect linguistic complexity. Previous studies have addressed the L1-related differences in various aspects of L2 writing, such as information structure, syntactic patterns, and lexical style, highlighting the necessity for a more careful examination of potential L1 effects in L2 complexity research (Lu & Ai, 2015). We believe that the findings of the present study will provide valuable insights into assessing L2 writing performances and help teachers implement effective pedagogical interventions by demonstrating how linguistic complexity differs between different types of English users and how L1 backgrounds affect linguistic complexity.

2 | LITERATURE REVIEW

2.1 Defining linguistic complexity

In L2 research, the term complexity comprises various levels and dimensions, thus lacking a clear-cut definition (Bulté & Housen, 2012; Pallotti, 2015). To uncover the multidimensional nature of L2 complexity, Bulté and Housen (2012) classified L2 complexity into relative and absolute complexity. Relative complexity is a subjective and agent-related concept, which concerns the cognitive difficulty of processing perceived by language users (Miestamo et al., 2008). By contrast, absolute complexity is an objective notion and is defined as the formal properties inherent in a linguistic system (Bulté & Housen, 2012). To elaborately gauge L2 learners' performance, the general notion of absolute complexity was further divided into three components: propositional complexity, discourse-interactional complexity, and linguistic complexity.

In the current study, we focus on linguistic complexity, which refers to the absolute, objective, and quantitative properties of language units, features, and (sub)systems (Bulté & Roothooft, 2020). More specifically, we propose to use Kolmogorov complexity, which is defined as the length of the shortest description that can reproduce the sample texts (Juola, 2008; Li et al., 2004).

2.2 | Metrics of linguistic complexity in second language research

Over the past decades, linguistic complexity has been widely operationalized using a variety of indices, some of which have been demonstrated to be reliable in gauging L2 learners' language development and global proficiency. Precisely, concerning lexical complexity, lexical diversity metrics, including type-token ratio and Guiraud's index

(Bulté & Roothooft, 2020; De Clercq, 2015; Treffers-Daller et al., 2016) and more recent lexical sophistication metrics such as n-gram association strength (Kim et al., 2017; Kyle et al., 2017) are proved to be effective in differentiating proficiency levels. Regarding syntactic complexity, previous studies demonstrated that measures dealing with the average length of production units (e.g., sentences, clauses, or T-units) and subordination (e.g., dependent clauses per clause and dependent clauses per T-unit) increased with proficiency levels (Lu, 2011; Ortega, 2003; Ouyang et al., 2022). More recent studies have addressed the call for finer-grained indices such as clausal complexity and phrasal complexity metrics (Kyle & Crossley, 2017, 2018; Zhang & Lu, 2022) and proved their predictive power of proficiency.

By contrast, few studies have examined the relationship between morphological complexity and L2 proficiency (Brezina & Pallotti, 2016; De Clercq & Housen, 2019; Pallotti, 2015; Yoon, 2018). Yoon (2018) investigated the syntactic, lexical, and morphological dimensions of language development among learners of varying proficiency levels. Concerning morphological complexity, results showed that significant changes existed across proficiency levels, but no significant difference was observed between adjacent proficiency levels. Yoon's (2018) study is significant as it points out the potential applicability of morphological diversity in predicting learner proficiency. However, the morphological complexity index (MCI) adopted by Yoon (2018) was later demonstrated to be less predictable in gauging learner proficiency than the Kolmogorov morphological complexity, which exhibited a larger effect size in explaining differences across proficiency levels (Wang, Wang, & Wang, 2022).

Moreover, Wang, Wang, and Wang (2022) found that Kolmogorov overall and syntactic complexity performed best in distinguishing L2 proficiency, as compared to traditional syntactic and morphological complexity metrics as well as fine-grained syntactic complexity metrics. This positions the Kolmogorov complexity as a robust tool, capable of predicting learner proficiency through its three layers: overall, morphological, and syntactic.

To summarize, the studies reviewed have provided various effective indices that can reliably uncover the relationship between linguistic complexity and L2 proficiency. However, most of these studies fixed their eyes on lexical or syntactic complexity, with limited empirical studies targeting morphological complexity. Furthermore, except for Kolmogorov complexity, none of these metrics can capture the complex and multidimensional nature of L2 complexity simultaneously. Therefore, we propose to use Kolmogorov complexity, an information-theoretic metric, which can address three sublevels of linguistic complexity (i.e., overall, syntactic, and morphological complexity) at the same time, thus serving as an effective complement to previous studies.

2.3 | Kolmogorov complexity

^₄ WILEY

Kolmogorov complexity is defined as the length of the shortest possible description to regenerate the running texts (Juola, 2008; Li et al., 2004). Due to some mathematical problems, it is difficult to calculate this complexity metric directly (Kolmogorov, 1968). Instead, we can approximately compute Kolmogorov complexity by using an entropy estimation approach with the help of file compression programs like gzip, whose algorithm is built on structural redundancies of the sample texts.

Linguistically speaking, Kolmogorov complexity metrics differ from traditional complexity metrics in that traditional ones tend to emphasize certain structural and grammatical features, such as dependent clauses and relative clauses. On the contrary, Kolmogorov complexity is not feature specific but holistic as it takes the whole structural complexity of running texts into consideration. In other words, Kolmogorov complexity is not relevant to the meaningcarrying grammatical features but emphasizes surface structural redundancy, which deals with the recurrence or repetition of orthographic character sequences within a text (Ehret, 2021).

Therefore, Kolmogorov complexity is insufficient in detecting the changes of specific linguistic features compared with traditional complexity metrics. However, considering the significant effect of Kolmogorov complexity metrics in differentiating learner proficiency (Wang et al., 2022), we propose that Kolmogorov complexity metrics may complement the traditional ones to depict the variations of linguistic complexity across speaker types and L1 back-grounds. Additionally, Kolmogorov complexity's ability to evaluate overall, syntactic, and morphological complexity

concurrently provides a comprehensive and multidimensional instrument for capturing the intricate nature of L2 complexity (Ehret & Szmrecsanyi, 2016).

Kolmogorov complexity was first utilized by Juola (1998) and afterward introduced in linguistic research. So far, it has been applied to investigate cross-linguistic complexity differences by analyzing parallel corpora that contain the original sample texts and their translations (Ehret & Szmrecsanyi, 2016; Juola, 2008; Sadeniemi et al., 2008).

More recently, Kolmogorov complexity has been proved to be applicable to non-parallel corpus data (Ehret, 2021; Ehret & Szmrecsanyi, 2019; Wang et al., 2022; Wang et al., 2022). For instance, based on naturalistic second language acquisition data, Ehret and Szmrecsanyi (2019) employed Kolmogorov complexity to investigate the relationship between the complexity of L2 English learners' essays and the amount of instruction received by the essay writers. Results showed that increased L2 instructional exposure predicts increased overall complexity and morphological complexity but decreased syntactic complexity. Using British National Corpus, Ehret (2021) explored complexity differences between written and spoken registers of British English. The findings indicated that these two registers have no absolute difference in complexity, and that the difference is gradual rather than rigid. To conclude, those studies have provided compelling evidence for the reliability and validity of Kolmogorov complexity in conducting linguistic research.

2.4 Linguistic complexity across L1 backgrounds

By virtue of complexity metrics, researchers have investigated the relationship between linguistic complexity and various task, context, or learner-related factors (e.g., Seyyedi et al., 2013; Tabari & Wang, 2022; Yoon, 2021). One specific learner-related factor that has received increasing attention is learners' L1 backgrounds.

For instance, Crossley and McNamara (2012) examined L2 English essays written by speakers with four L1 backgrounds and detected significant differences in syntactic complexity among these four groups. However, Crossley and McNamara (2012) only examined one syntactic complexity measure, that is, the mean number of words before the main verb. Lu and Ai (2015) extended the research scope by investigating the differences of 14 syntactic complexity indices in argumentative essays written by NS and EFL learners of 7 L1 backgrounds. Results indicated that despite having similar proficiency levels, learners with different L1 backgrounds might not develop in the same ways in some sublevels of syntactic complexity (e.g., degree of coordination and subordination). Additionally, Khushik and Huhta (2020) examined the syntactic complexity of writings produced by learners from Pakistan and Finland across three proficiency levels. They found that when proficiency was controlled, the most significant differences between learners across L1 backgrounds were found in the length-based metrics and phrasal density.

These studies contribute to our understanding of the effects of L1 backgrounds on linguistic complexity. However, to our knowledge, none of them has explored the differences in linguistic complexity between NS, ESL learners, and EFL learners. The study that most resembles our sphere of interest is that of Barrot and Gabinete (2021). They investigated whether there was a difference in the complexity, accuracy, and fluency in the argumentative writings of ESL and EFL learners. Specifically, Barrot and Gabinete (2021) found that essays written by ESL learners were significantly more complex than those of EFL learners regarding the mean length of clauses and the percentage of dependent clauses of all the clauses. However, Barrot and Gabinete (2021) merely investigated syntactic complexity and did not use NS as a baseline to measure the performances of EFL/ESL learners.

To these ends, based on 2272 essays collected from the International Corpus Network of Asian Learners of English (ICNALE)-Written, this study attempts to investigate the differences in linguistic complexity between the writings of NS, EFL learners, and ESL learners by using a novel information-theoretic Kolmogorov complexity at three complexity levels (i.e., overall, morphological, and syntactic complexity). Furthermore, we have investigated the effects of different L1 backgrounds on linguistic complexity. The research questions are as follows:

1. Are there any significant differences between essays written by different types of English speakers (i.e., ESL, EFL, and NS) in terms of overall, morphological, and syntactic complexity, respectively?

• WILEY-

2. Are there any significant differences in the linguistic complexity of essays written by learners with different L1 backgrounds, and if yes, what are these differences?

3 | METHODOLOGY

3.1 | Corpus data

We used the ICNALE-Written (Ishikawa, 2011) as our corpus. The reason for choosing this corpus is threefold.

First, the ICNALE-Written, comprising 5600 argumentative written essays and amounting to 1.3 million tokens, is the largest international learner corpus focusing on Asian learners' English. Specifically, this corpus contains writings produced by 2800 intermediate to advanced learners, who are categorized into four Common European Framework of Reference (CEFR)-linked proficiency levels (i.e., A2_0, B1_1, B1_2, and B2_0) based on the scores they received on the TOEIC, TOEFL, IELTS, or the English vocabulary size test (Nation & Beglar, 2007). In addition, the ICNALE-Written includes separate data from four ESL countries/regions (i.e., Hong Kong, Pakistan, the Philippines, and Singapore) and six EFL countries/regions (i.e., China, Indonesia, Japan, Korea, Taiwan, and Thailand), thus providing us with a large number of argumentative writing samples for analysis across various countries/regions.

Second, all the writing essays along with authors' relevant metadata (e.g., age, English type, and English level) can be freely accessed at the ICNALE homepage (http://language.sakura.ne.jp/icnale/). Based on the metadata, subcorpora could be further extracted to accomplish our research objective. In this study, we specifically extracted essays with B1_2 level (an upper-intermediate CEFR-linked level) as the learner corpus data since this category provides a sufficient number of essays and tokens across Asian countries and regions (Barrot & Gabinete, 2021). In contrast, the essays at other levels are either unavailable or limited in size for given countries.

Third, the ICNALE-Written rigidly controls the prompts and tasks of the writing process, such as the writing topics, the time for writing an essay, and the length of an essay, thus guaranteeing a reliable source for our study. More precisely, each learner contributes two essays on the given topics (i.e., a part-time job for college students and nonsmoking at restaurants) for about 200 to 300 tokens within 20 to 40 min (Ishikawa, 2011). The statistical overview of the final data used in the present study is shown in Table 1.

It is necessary to note that we integrated two paired essays written by the same learner due to the influence of text length on the calculation of Kolmogorov complexity. Specifically, complexity measurements are more robust and representative if they are based on larger texts (Ehret & Szmrecsanyi, 2016). Furthermore, as these texts share similar task complexity characteristics (e.g., writing genre, time pressure, and production mode), we propose that there should be no discernible differences between the two samples based on the trade-off hypothesis and cognition hypothesis (Foster & Skehan, 1996; Robinson, 2001).

3.2 | The calculation of Kolmogorov complexity

The Kolmogorov complexity of a text can be measured by the length of the shortest description to restate it using file compression programs (Li et al., 2004; Juola, 2008). Specifically, texts that can be compressed more efficiently hold a lower Kolmogorov complexity, whereas less compressible texts possess a higher Kolmogorov complexity (Ehret & Szmrecsanyi, 2019).

The following two strings are used to elaborate the calculation of Kolmogorov complexity. Although both Strings A and B contain 8 characters, String A can be compressed as 4 times *ng*, comprising 4 characters, while String B is uncompressible as it lacks any recurring pattern. Therefore, concerning Kolmogorov complexity, String A is less complex than String B.

| | | | Words per text | | Sentences | |
|----------|-------------------|--------------|----------------|-------|-----------|-------------|
| Category | Countries/regions | No. of texts | Mean | SD | per text | Total words |
| NS | | 200 | 449.80 | 42.37 | 18.6 | 89,959 |
| EFL | China | 105 | 490.17 | 63.92 | 30.4 | 51,468 |
| | Indonesia | 83 | 473.66 | 50.90 | 28.6 | 39,314 |
| | Japan | 49 | 451.67 | 41.38 | 31.5 | 22,132 |
| | South Korea | 88 | 451.24 | 52.42 | 33.5 | 39,709 |
| | Taiwan | 61 | 452.03 | 43.03 | 27.0 | 27,574 |
| | Thailand | 100 | 472.55 | 58.44 | 27.4 | 47,255 |
| | | 486 | 468.01 | 55.94 | 29.7 | 227,452 |
| ESL | Hong Kong, China | 52 | 471.88 | 62.41 | 27.3 | 24,538 |
| | Pakistan | 88 | 471.10 | 64.32 | 29.9 | 41,457 |
| | The Philippines | 176 | 456.02 | 49.73 | 25.3 | 80,260 |
| | Singapore | 134 | 491.97 | 56.63 | 22.8 | 65,924 |
| | | 450 | 471.51 | 58.11 | 25.7 | 212,179 |

TABLE 1 Descriptive statistics of the data used in the study.

A. ngngngng (8 characters) – $4 \times$ ng (4 characters)

B. gald9G5x (8 characters) – gald9G5x (8 characters)

To facilitate the application of Kolmogorov complexity in linguistic research, Ehret (2017) introduced a compression technique based on Juola's (2008) algorithm. In the current study, we replicated the pioneering work of Ehret (2017 pp. 43–82) regarding the newly proposed compression technique and the steps for calculating Kolmogorov complexity. To be specific, gzip (GNU zip, Version 1.11, https://ftp.gnu.org/gnu/gzip/) was employed to estimate the Kolmogorov complexity of each text at the overall, syntactic, and morphological level. Additionally, the scripts for implementing this compression technique can be accessed at GitHub: https://github.com/katehret/measuring-language-complexity.

3.2.1 | Overall complexity

The overall complexity of a text is congruent with Miestamo et al.'s (2008) concept of global complexity, which includes the complexity of all levels of a language, thus addressing the entire structural complexity of a text.

To compute the overall complexity, we first assessed the file size (in bytes) before and after compression for each text. Subsequently, we employed a linear regression analysis with the uncompressed file size as the independent variable and compressed file size as the dependent variable, thus eliminating the correlation between them. This step yields the adjusted overall complexity scores (i.e., regression residuals) of sample texts: the higher the score, the higher the overall linguistic complexity is.

3.2.2 | Morphological complexity

To assess the morphological complexity, we first randomly deleted 10% (a conventional percentage employed in previous studies; see Ehret & Taboada, 2021; Juola, 1998; Sadeniemi et al., 2008) of the characters before compressing the running texts. Then the distorted texts were compressed to identify how well or badly the compression technique

deals with the distortion. Equation (1) displays the algorithm of morphological complexity.

Morphological complexity score
$$= -\frac{m}{c}$$
 (1)

In Equation (1), *m* refers to the compressed file size after morphological distortion, and *c* is the original compressed file size. Considering that morphologically complex texts tend to contain a relatively larger number of word forms, they will be less affected by the distortion process than morphologically simple texts, for which distortion may negatively affect their compressibility. Therefore, comparatively bad compression ratios after morphological distortion indicate low morphological complexity, and vice versa.

3.2.3 | Syntactic complexity

To calculate syntactic complexity, we randomly deleted 10% of all word tokens in each text. Then we compressed the distorted texts and obtained the syntactic complexity scores of given texts according to Equation (2).

Syntactic complexity score
$$=\frac{s}{c}$$
 (2)

In Equation (2), s stands for the compressed file size after syntactic distortion, and c represents the file size before distortion. One point worth noting is that in the current study, syntactic complexity is understood as word order rigidity (Bakker, 1998): rigid word order suggests syntactically complex texts, while free word order is characteristic of syntactically simple texts. Syntactic distortion, then, disrupts word order regularities, resulting in random noises. Syntactically complex texts are greatly influenced, and their compression efficiency is compromised; syntactically simple texts, in contrast, are less affected due to a lack of syntactic interdependencies that could be compromised. Therefore, relatively bad compression ratios after syntactic distortion indicate high syntactic complexity.

This seems to be counterintuitive as one might intuitively assume that free word order, characterized by lower predictability, should be more complex than rigid word order. Nonetheless, we should bear in mind that, Kolmogorov syntactic complexity is computed indirectly since we measure to what extent distortion will affect the predictability of a text. If a text becomes less predictable after distortion, then it can be considered as syntactically complex. In this regard, rigid word order is regarded as Kolmogorov complex from a technical point of view (Ehret & Szmrecsanyi, 2019, p. 28).

3.3 | Data processing

In this section, we depicted the procedures of data processing. All the procedures were carried out using homemade scripts in R, a programming language for data processing and statistical analysis.

3.3.1 | Data collection

We extracted all the essays with B1_2 level (the upper-intermediate level) as well as the essays written by native speakers and combined the two paired essays by the same learner. This step produced 1136 texts, which were then used as the data.

3.3.2 | Data cleaning

All running texts were lowercased, and non-alphabetical characters were removed (e.g., numbers, UTF-8 characters, and corpus markups), along with punctuation (e.g., dashes, commas, and hyphens). We did this because punctuations and non-alphabetical characters would compromise the compressibility of texts and thus increase their complexity. We retained the full stops and replaced other end-of-sentence markings (e.g., semicolons, exclamation marks, and question marks) with full stops. This is because full stops serving as the end markers of sentences are used to determine the linguistic units of random sampling in Kolmogorov complexity calculation. In addition, we manually checked all the possible mistakes resulting from the deletion of numbers and punctuations.

3.3.3 | Kolmogorov complexity calculation

To generate a statistically robust result, we repeated the distortion and compression process for each text for 500 times. In each iteration, we employed random sampling, that is, randomly selected five sentences per text. We did this because random sampling keeps sample size constant, thus ensuring the comparability of linguistic metrics among texts of different sizes.

To measure the overall Kolmogorov complexity, we calculated the mean file sizes before and after compression across all iterations. Subsequently, a linear regression was performed, and the adjusted overall complexity scores for each text were calculated. For the morphological and syntactic complexity, we first calculated their scores for each text file in each iteration. Then, the average morphological and syntactic complexity scores for each text were computed across all iterations, respectively.

3.4 | Statistical analyses

For the first research question, to determine the complexity differences between NS, ESL learners, and EFL learners, a non-parametric Kruskal–Wallis test was performed on each of the three linguistic measures (i.e., overall, morphological, and syntactic complexity) across the NS, EFL, and ESL groups. Kruskal–Wallis test was used because the Shapiro–Wilk test and Q–Q plots of the three complexity measures showed that none of them were normally distributed. Then, paired comparisons with Bonferroni correction were carried out to determine whether the differences were significant in the linguistic measures between every two groups (i.e., NS and ESL; NS and EFL; EFL and ESL).

For the second research question, we first conducted the Kruskal–Wallis test to examine the complexity differences between NS and the four ESL countries/regions. Then, Dunnett's *t*-tests with Bonferroni correction were employed to determine whether there is a significant difference between NS and any of the four ESL countries/regions. The same statistical tests (i.e., Kruskal–Wallis tests and the following Dunnett's *t*-tests) were adopted to determine the complexity differences between NS and the six EFL countries/regions. In addition, we fitted simple linear regression models to each linguistic complexity measure to further examine the predictive power of L1 backgrounds and speaker types on complexity scores.

4 | RESULTS

This section first reports the linguistic complexity differences between the NS, ESL, and EFL groups, which were each regarded as a whole group. We then compared each of the four ESL countries/regions against the NS group, and also each of the six EFL countries/regions against the NS group to further examine the effects of L1 backgrounds on linguistic complexity.

| language (ESL), and English as a foreign language (EFL) groups. | | | | | | | | | | |
|---|--------|--------|--------|-------|-------|-------|--|--|--|--|
| | Mean | | | SD | | | | | | |
| Metrics | NS | ESL | EFL | NS | ESL | EFL | | | | |
| Overall | 6.554 | 0.139 | -2.826 | 8.019 | 9.348 | 8.375 | | | | |
| Syntactic | 0.924 | 0.927 | 0.930 | 0.003 | 0.003 | 0.004 | | | | |
| Morphological | -0.964 | -0.964 | -0.963 | 0.009 | 0.010 | 0.011 | | | | |

TABLE 2 Descriptive statistics of complexity measures across the native speakers (NS), English as a second language (ESL), and English as a foreign language (EFL) groups.

Abbreviations: Overall, overall complexity; Syntactic, syntactic complexity; Morphological, morphological complexity.



FIGURE 1 Paired comparisons across the native speakers (NS), English as a second language (ESL), and English as a foreign language (EFL) groups on overall complexity and syntactic complexity. [Color figure can be viewed at wileyonlinelibrary.com]

4.1 | Linguistic complexity across the NS, ESL, and EFL groups

The descriptive statistics of the linguistic measures across the NS, ESL, and EFL groups are presented in Table 2. It is shown that the mean value of overall complexity followed the order of NS > ESL > EFL from the most complex to the least, while the order was reversed for syntactic complexity: NS < ESL < EFL. Regarding morphological complexity, the NS and ESL groups possessed almost the same mean value, but what is surprising is that EFL was the most complex: EFL > NS = ESL.

To determine whether there were any significant differences in linguistic complexity across the NS, ESL, and EFL groups, we conducted the Kruskal–Wallis tests. The results showed that speaker types had a significant effect on overall complexity (H(2) = 203.93, p = 0.000, eta2[H] = 0.18) and syntactic complexity (H(2) = 357.15, p = 0.000, eta2[H] = 0.31) with a large effect size. However, no significant difference was found across the NS, ESL, and EFL groups for morphological complexity.

To further determine the differences in linguistic complexity between any two groups (ESL and EFL; ESL and NS; EFL and NS), we performed paired comparisons with Bonferroni correction. The results with Bonferroni-corrected *p* values are plotted in Figure 1.



FIGURE 2 Comparison between native speakers and each individual English as a second language country/region. *Note:* * < = 0.05; ** < = 0.01; *** < = 0.001; *** < = 0.001; ns, no significant difference. [Color figure can be viewed at wileyonlinelibrary.com]

As shown in Figure 1, there was a significant difference between any two groups for both overall complexity and syntactic complexity. Concerning the medians of complexity scores, the NS group produced essays with the highest overall complexity but the lowest syntactic complexity. In contrast, the essays written by the EFL group exhibited the highest syntactic complexity but the lowest overall complexity.

4.2 | Linguistic complexity across various L1 backgrounds

The Kruskal–Wallis tests revealed significant differences among the NS group and the four ESL countries for all three complexity metrics. In addition, a large effect size was found both on overall complexity (H(4) = 124.10, p = 0.000, eta2[H] = 0.19) and syntactic complexity (H(4) = 176.80, p = 0.000, eta2[H] = 0.27), whereas small on morphological complexity (H(4) = 22.46, p = 0.000, eta2[H] = 0.03).

Subsequently, we employed follow-up paired comparisons to determine whether there were significant differences between NS and any of the four ESL countries/regions on the complexity metrics. Figure 2 shows the boxplots of complexity scores of NS and the four ESL countries/regions across the overall, syntactic, and morphological levels. Note that the median scores and the results of the pairwise comparisons between NS and any of the four ESL countries/regions were provided. The red dotted lines indicate the median scores of each complexity metric for NS.

All four ESL countries/regions had significantly lower overall complexity than NS, and all ESL countries/regions had significantly higher syntactic complexity than NS. Interestingly, Singapore ranked closest to NS, while Pakistan differed most from it in terms of overall and syntactic complexity. Regarding morphological complexity, Pakistan was significantly higher than NS, whereas the Philippines, Singapore, and Hong Kong showed no significant difference from NS.

Kruskal–Wallis tests showed significant effects of the L1 backgrounds of EFL learners on all three complexity measures: overall complexity (H(6) = 258.30, p = 0.000, eta2[H] = 0.37), syntactic complexity (H(6) = 325.01, p = 0.000, eta2[H] = 0.47), and morphological complexity (H(6) = 81.50, p = 0.000, eta2[H] = 0.11). The eta-squared estimates of overall complexity and syntactic complexity both exceeded 0.14, indicating a large effect size.



FIGURE 3 Comparison between native speakers and each individual English as a foreign language country/region. *Note*: ***< = 0.0001; ****< = 0.0001; ns, no significant difference. [Color figure can be viewed at wileyonlinelibrary.com]

Follow-up paired comparisons were adopted to determine whether there were significant differences between NS and any of the six EFL countries/regions on the overall, syntactic, and morphological complexity. As shown in Figure 3, the median scores of all the six EFL countries/regions were significantly lower than NS in terms of overall complexity, but significantly higher in terms of syntactic complexity. In addition, among the six EFL countries/regions, China was the closest to NS in both overall and syntactic complexity. Concerning morphological complexity, all countries/regions showed no significant difference with NS except for China, which was the most morphologically complex among the EFL group, surpassing even the NS group.

To further determine whether the L1 backgrounds and English speaker types (i.e., NS, EFL, and ESL) significantly predict linguistic complexity scores, we fitted simple linear regression models across each complexity measure.

Results showed that both the L1 backgrounds and English speaker types significantly predicted the overall complexity (English speaker types: $R^2 = 0.125$, F(2, 1133) = 82.2, p = < 0.001; L1 backgrounds: $R^2 = 0.189$, F(10, 1125) = 27.4, p = < 0.001) and syntactic complexity (English speaker types: $R^2 = 0.293$, F(2, 1133) = 235.8, p = < 0.001; L1 backgrounds: $R^2 = 0.388$, F(10, 1125) = 72.9, p = < 0.001). As for morphological complexity, only the L1 backgrounds ($R^2 = 0.076$, F(10, 1125) = 10.4, p = < 0.001) serves as a significant predictor.

In addition, the L1 backgrounds were generally a more powerful predictor than English speaker types, as the adjusted R^2 values for L1 backgrounds were larger than those of English speaker types in all three complexity metrics. Results also revealed that the effect of L1 backgrounds caused the greatest variation in syntactic complexity ($R^2 = 0.388$) followed by overall complexity ($R^2 = 0.189$) and morphological complexity ($R^2 = 0.076$), suggesting that syntactic complexity is the most discriminatory metric among L1 backgrounds.

5 DISCUSSION

12

Based on the ICNALE-Written, the present study first investigated the differences in linguistic complexity across the NS, EFL, and ESL groups. In addition, this study identified the influence of L1 backgrounds on linguistic complexity. To

our knowledge, this is the first study that utilizes an information-theoretic metric (here Kolmogorov complexity) to explore linguistic complexity differences between NS, EFL learners, and ESL learners. Some possible explanations and implications of our findings will be discussed in detail below.

5.1 | Effects of English speaker types on linguistic complexity

Regarding the first research question, the current study found that when NS, EFL, and ESL were each regarded as a whole group, significant differences were detected between any two groups regarding overall and syntactic complexity. However, there was no significant difference across the NS, EFL, and ESL groups concerning morphological complexity.

Specifically, as regards overall complexity, results showed that essays written by NS were the most complex, followed, in decreasing order, by ESL and EFL learners. Such a result may relate to varying degrees of exposure to English. Specifically, for native speakers, exposure to the English language can occur anywhere and at any time. In ESL context, language learning occurs as a side effect of being involved in daily activities including living and working; hence, learners have frequent exposure to English (Pecorari, 2018). By contrast, in EFL context, English is principally acquired in the classroom via relevant class-related activities (Nayar, 1997), thus resulting in relatively limited exposure for learners. Therefore, the group with greater exposure to English (NS > ESL > EFL) is expected to produce more Kolmogorov complex essays.

When contrasted with overall complexity, the syntactic complexity scores exhibited a reverse ranking: NS < ESL < EFL. Syntactic complexity in the present study is associated with word order rigidity: rigid word order is considered as complex, whereas varied word order is treated as simple. Thus, our study indicates that EFL learners produce the writings with more rigid word order patterns. Considering the language input available to learners, this may result from the fact that EFL students acquire English primarily in the classroom and receive classroom-based and mostly form-focused instructions (Pecorari, 2018). As a result, extensive attention has been paid to the practice of grammatical patterns. Furthermore, students are encouraged to produce writings with memorized phrases and complex clauses, because they are made aware that sentences with sophisticated structures earn them higher marks. Therefore, EFL learners often use more complex syntax than native speakers, even though it is not necessarily appropriate.

Nevertheless, our findings are not consistent with those reported by Barrot and Gabinete (2021), who found that ESL learners produce more syntactically complex writings than EFL learners. This discrepancy might be due to the different linguistic aspects targeted by the complexity metrics employed. Specifically, complexity metrics used in Barrot and Gabinete (2021) address clausal subordination and phrasal complexification, while Kolmogorov syntactic complexity adopted in our study is related to word order rigidity. In this respect, Kolmogorov complexity may reveal unique features that may not be captured by other metrics, thus complementing the previous literature concerning L1-related complexity differences.

Another important finding is that no significant difference was found among the NS, ESL, and EFL groups on morphological complexity, indicating that essays written by NS did not exhibit more morphological forms than texts produced by EFL and ESL learners at the upper-intermediate level. These results further support the findings of Brezina and Pallotti (2016), who found that morphological complexity remains constant across native English speakers and non-native proficient English learners. Possible explanations might lie in the characteristics of the target language and learners' proficiency levels (DeKeyser, 2016). Specifically, compared with languages like Italian that enjoy a rich array of inflectional endings, English possesses much fewer inflectional resources, which would be much easier for learners to acquire. Thus, learners' inflectional diversity will remain constant once they reach a relatively high proficiency level.

5.2 | Effects of L1 backgrounds on linguistic complexity

The second research question is to determine the differences in linguistic complexity within different L1 backgrounds. Our results showed that complexity variations exist among the four ESL countries and among the six EFL countries across the three complexity metrics, as compared with NS. Considering that all the texts were extracted from the same proficiency level, the intergroup variations in linguistic complexity may be accounted for by learners' L1 backgrounds.

These results indicate the necessity of closer scrutiny of the effects of L1 backgrounds on linguistic complexity, which is in line with the findings of previous studies (Barrot & Gabinete, 2021; Ehret & Szmrecsanyi, 2019; Lu & Ai, 2015; Ortega, 2015). As argued by Lu and Ai (2015), learners with different L1 backgrounds, even for those at the same or comparable proficiency levels, might not develop in the same ways in syntactic complexity. Consequently, they proposed that the L1 backgrounds of L2 writers cannot be ignored, when examining syntactic complexity in L2 writing.

In addition, our results showed that in comparison to English speaker types, L1 background was a more highly predictive variable on all three complexity metrics, indicating that teachers should pay more attention to learners' L1 backgrounds when attempting to incorporate pedagogical interventions into the L2 writing process. These results are consistent with those of Ehret and Szmrecsanyi (2019). They propose that L1 backgrounds serves as a strong predictor of overall complexity. Specifically, Ehret and Szmrecsanyi (2019) found that German learners of English produced more Kolmogorov complex essays in terms of overall and morphological complexity than others from French, Italian, or Spanish. It is noteworthy that our study used a learner corpus covering a much wider range of L1 backgrounds compared to that of Ehret and Szmrecsanyi (2019).

Moreover, among the four ESL countries, Singapore is the closest one to NS, while China is the closest to NS among the six EFL countries, either for overall complexity or syntactic complexity. The similar complexity level to NS shown by Singapore might be attributed to its English-based bilingual education policies and various campaigns promoting the use of English, such as the Speak Good English Movement launched in 2000. These practices may have facilitated Singapore's integration into global communication and contributed to maintaining its reputation as a leading financial and commercial hub (Tang, 2020).

As for China, in accordance with the present result, previous studies have demonstrated that China outperforms even some ESL countries on given complexity metrics (e.g., phrasal complexity and sentential coordination) (Lu & Ai, 2015). This result may be explained by the fact that Chinese people have made great strides in their English language skills, as China has emerged as a major economic powerhouse. China attaches a great deal of importance to English study, since it contributes to one's future career options and the pursuit of further education. It is believed that people who are proficient in English are more competitive than their peers.

6 CONCLUSIONS

Our study aims to explore how linguistic complexity differs between NS, ESL learners, and EFL learners by using a holistic and information-theoretic Kolmogorov complexity. Based on 2272 written essays in the ICNALE-Written, we investigated the variations of linguistic complexity (here overall, morphological, and syntactic complexity) between the NS, EFL, and ESL groups. In addition, we explored the effects of L1 backgrounds on linguistic complexity by investigating the complexity differences between NS and four ESL countries/regions, and between NS and six EFL countries/regions.

As for the first research question, results showed that overall, NS produced the most complex essays, followed in decreasing order by ESL and EFL, while the syntactic complexity exhibited a reverse order. These results may be attributed to the diverse degree and types of exposure to English across the three groups. Concerning morphological complexity, no significant difference was found between the NS, ESL, and EFL groups. As for the second research question, we found that with respect to overall complexity and syntactic complexity, Singapore and China displayed

WILEY <u>¹⁵</u>

a similar complexity level to the NS group, possibly due to their heavy emphasis on English education and diverse practices to enhance the use of English. These results reveal a strong effect of L1 backgrounds on linguistic complexity.

The findings reported here shed new light on the linguistic complexity studies of L2 learners, which has several implications for both teachers and researchers alike. Pedagogically speaking, teacher awareness of L1-related differences in linguistic complexity in L2 writing is crucial for implementing effective L2 writing strategies tailored to learners with diverse L1 backgrounds. For researchers, our results suggest that they should be aware of why and when learners' L1 backgrounds need to be considered in their research design.

In addition, our study provides methodological contributions to the current complexity literature. Specifically, Kolmogorov complexity is more global and comprehensive than traditional measures that only depict several specific linguistic features, or only cover certain aspects of linguistic levels (e.g., grammatical and lexical levels). Thus, this complexity measure is well suited for capturing the complex multidimensional nature of L2 complexity (Ehret & Szmrecsanyi, 2016). Additionally, Kolmogorov complexity as a holistic and quantitative measure of text complexity is both more economical to obtain and arguably more objective than, for example, subjective complexity ratings of learner texts by expert evaluators.

The major limitation of this study is that we have only selected argumentative essays written by learners of 10 upper-intermediate level Asian countries/regions as the data. In the future, our findings could be enhanced by using other corpora on learner English that contain more countries and genres, or by examining other levels of students' essays, thus further evaluating the validity and reliability of this methodology. Furthermore, we combined two essays by the same learner considering the effect of text length on the calculation of Kolmogorov complexity. However, this may weaken the study's validity as topic influences linguistic complexity (Yoon, 2018); hence, future studies are warranted to examine the effect of topics on Kolmogorov complexity using a larger corpus.

ACKNOWLEDGMENT

This work was supported by the National Social Science Foundation of China (No. 17BYY115).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in International Corpus Network of Asian Learners of English at http://language.sakura.ne.jp/icnale. These data were derived from the following resources available in the public domain: Written Essays, http://language.sakura.ne.jp/icnale/download.html

ORCID

Li Wang D https://orcid.org/0000-0002-3028-1347

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/ijal.12526

REFERENCES

- Bakker, D. (1998). Flexibility and consistency in word order patterns in the languages of Europe. In A. Siewierska (Ed.), Constituent order in the languages of Europe (pp. 383–420). De Gruyter Mouton. https://doi.org/10.1515/9783110812206. 383
- Barrot, J., & Gabinete, M. (2021). Complexity, accuracy, and fluency in the argumentative writing of ESL and EFL learners. International Review of Applied Linguistics in Language Teaching, 59(2), 209–232.
- Biber, D., Gray, B., Staples, S., & Egbert, J. (2020). Investigating grammatical complexity in L2 English writing research: Linguistic description versus predictive measurement. *Journal of English for Academic Purposes*, 46, 100869. https://doi.org/10. 1016/j.jeap.2020.100869

- Brezina, V., & Pallotti, G. (2016). Morphological complexity in written L2 texts. Second Language Research, 35(1), 99-119. https://doi.org/10.1177/0267658316643125
- Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA (pp. 23–46). John Benjamins.
- Bulté, B., & Roothooft, H. (2020). Investigating the interrelationship between rated L2 proficiency and linguistic complexity in L2 speech. System, 91, 102246. https://doi.org/10.1016/j.system.2020.102246
- Crossley, S. A., & McNamara, D. S. (2012). Detecting the first language of second language writers using automated indices of cohesion, lexical sophistication, syntactic complexity and conceptual knowledge. In S. Jarvis & S. Crossley (Eds.), Approaching language transfer through text classification: Explorations in the detection-based approach. Multilingual Matters.
- De Clercq, B. (2015). The development of lexical complexity in second language acquisition: A cross-linguistic study of L2 French and English. EUROSLA Yearbook, 15(1), 69–94. https://doi.org/10.1075/eurosla.15.03dec
- De Clercq, B., & Housen, A. (2019). The development of morphological complexity: A cross-linguistic study of L2 French and English. Second Language Research, 35(1), 71–97. https://doi.org/10.1177/0267658316674506
- DeKeyser, R. (2016). Of moving targets and chameleons. Studies in Second Language Acquisition, 38(2), 353–363. https://doi. org/10.1017/s0272263116000024
- Ehret, K. (2017). An information-theoretic approach to language complexity: Variation in naturalistic corpora [Doctoral Dissertation, University of Freiburg].
- Ehret, K. (2021). An information-theoretic view on language complexity and register variation: Compressing naturalistic corpus data. Corpus Linguistics and Linguistic Theory, 17(2), 383-410. https://doi.org/10.1515/cllt-2018-0033
- Ehret, K., & Szmrecsanyi, B. (2016). An information-theoretic approach to assess linguistic complexity. In R. Baechler & G. Seiler (Eds.), Complexity, isolation, and variation (pp. 71-94). De Gruyter.
- Ehret, K., & Szmrecsanyi, B. (2019). Compressing learner language: An information-theoretic measure of complexity in SLA production data. Second Language Research, 35(1), 23-45. https://doi.org/10.1177/0267658316669559
- Ehret, K., & Taboada, M. (2021). The interplay of complexity and subjectivity in opinionated discourse. Discourse Studies, 23(2), 141-165. https://doi.org/10.1177/1461445620966923
- Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. Studies in Second Language Acquisition, 18(3), 299-323. https://doi.org/10.1017/s0272263100015047
- Ishikawa, S. (2011). A new horizon in learner corpus studies: The aim of the ICNALE Project. In G. Weir, S. Ishikawa, & K. Poonpon (Eds.), Corpora and language technologies in teaching, learning and research (pp. 3–11). University Of Strathclyde Publishing.
- Jiang, J., Bi, P., Xie, N., & Liu, H. (2023). Phraseological complexity and low- and intermediate-level L2 learners' writing quality. International Review of Applied Linguistics in Language Teaching, 61(3), 765–790. https://doi.org/10.1515/iral-2019-0147
- Juola, P. (1998). Measuring linguistic complexity: The morphological tier. Journal of Quantitative Linguistics, 5(3), 206–213. https://doi.org/10.1080/09296179808590128
- Juola, P. (2008). Assessing linguistic complexity. In M. Miestamo, K. Sinnemäki, & F. Karlsson (Eds.), Language complexity: Typology, contact, change (pp. 89–108). John Benjamins. https://doi.org/10.1075/slcs.94.07juo
- Khushik, G. A., & Huhta, A. (2020). Investigating syntactic complexity in EFL learners' writing across common European framework of reference levels A1, A2, and B1. Applied Linguistics, 41(4), 506-532. https://doi.org/10.1093/applin/amy064
- Kim, M., Crossley, S., & Kyle, K. (2017). Lexical Sophistication as a Multidimensional Phenomenon: Relations to Second Language Lexical Proficiency, Development, and Writing Quality. The Modern Language Journal, 102(1), 120-141. https://doi. org/10.1111/modl.12447
- Kolmogorov, A. N. (1968). Three approaches to the quantitative definition of information. International Journal of Computer Mathematics, 2(1-4), 157-168. https://doi.org/10.1080/00207166808803030
- Kyle, K., & Crossley, S. (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. Language Testing, 34(4), 513-535. https://doi.org/10.1177/0265532217712554
- Kyle, K., & Crossley, S. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. The Modern Language Journal, 102(2), 333–349. https://doi.org/10.1111/modl.12468
- Kyle, K., Crossley, S., & Berger, C. (2017). The tool for the automatic analysis of lexical sophistication (TAALES): Version 2.0. Behavior Research Methods, 50(3), 1030-1046. https://doi.org/10.3758/s13428-017-0924-4
- Li, M., Chen, X., Li, X., Ma, B., & Vitanyi, P. M. B. (2004). The similarity metric. IEEE Transactions on Information Theory, 50(12), 3250-3264. https://doi.org/10.1109/tit.2004.838101
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. International Journal of Corpus Linguistics, 15(4), 474-496. https://doi.org/10.1075/ijcl.15.4.02lu
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. TESOL Quarterly, 45(1), 36-62. https://doi.org/10.5054/tq.2011.240859
- Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. Journal of Second Language Writing, 29, 16-27. https://doi.org/10.1016/j.jslw.2015.06.003

- Martínez-Adrián, M., & Nieva-Marroquín, M. (2023). Exploring the pseudo-longitudinal development of specific morphosyntactic features and syntactic complexity in CLIL young learners. *International Review of Applied Linguistics in Language Teaching*, 61(3), 1197–1223. https://doi.org/10.1515/iral-2021-0137
- Miestamo, M., Sinnemaki, K., & Karlsson, F. (2008). Language complexity: Typology, contact, change. John Benjamins Publication Company.
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. The Language Teacher, 31, 9-13.
- Nayar, P. B. (1997). ESL/EFL dichotomy today: Language politics or pragmatics? TESOL Quarterly, 31(1), 9. https://doi.org/10. 2307/3587973
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492–518. https://doi.org/10.1093/applin/24.4.492
- Ortega, L. (2015). Syntactic complexity in L2 writing: Progress and expansion. *Journal of Second Language Writing*, 29, 82–94. https://doi.org/10.1016/j.jslw.2015.06.008
- Ouyang, J., Jiang, J., & Liu, H. (2022). Dependency distance measures in assessing L2 writing proficiency. Assessing Writing, 51, 100603. https://doi.org/10.1016/j.asw.2021.100603
- Pallotti, G. (2015). A simple view of linguistic complexity. Second Language Research, 31(1), 117–134. https://doi.org/10.1177/0267658314536435
- Pecorari, D. (2018). English as a foreign language (EFL) versus English as a second language (ESL) writing. In J. I. Liontas (Ed.), The TESOL encyclopedia of English language teaching (pp. 1–6). Wiley Online Library.
- Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. Applied Linguistics, 22(1), 27–57. https://doi.org/10.1093/applin/22.1.27
- Sadeniemi, M., Kettunen, K., Lindh-Knuutila, T., & Honkela, T. (2008). Complexity of European Union languages: A comparative approach. Journal of Quantitative Linguistics, 15(2), 185–211. https://doi.org/10.1080/09296170801961843
- Seyyedi, K., Ismail, S. A. M. M., Orang, M., & Nejad, M. S. (2013). The effect of pre-task planning time on L2 learners' narrative writing performance. *English Language Teaching*, 6(12), 1–10. https://doi.org/10.5539/elt.v6n12p1
- Tabari, M. A., & Wang, Y. (2022). Assessing linguistic complexity features in L2 writing: Understanding effects of topic familiarity and strategic planning within the realm of task readiness. Assessing Writing, 52, 100605. https://doi.org/10.1016/j.asw. 2022.100605
- Tang, H. K. (2020). Linguistic landscaping in Singapore: Multilingualism or the dominance of English and its dual identity in the local linguistic ecology? International Journal of Multilingualism, 17(2), 152–173. https://doi.org/10.1080/14790718.2018. 1467422
- Treffers-Daller, J., Parslow, P., & Williams, S. (2016). Back to basics: How measures of lexical diversity can help discriminate between CEFR Levels. Applied Linguistics, 39(3), 203–327. https://doi.org/10.1093/applin/amw009
- Wang, G., Wang, H., Sun, X., Wang, N., & Wang, L. (2022). Linguistic complexity in scientific writing: A large-scale diachronic study from 1821 to 1920. *Scientometrics*, 128(1), 441–460. https://doi.org/10.1007/s11192-022-04550-z
- Wang, G., Wang, H., & Wang, L. (2022). Kolmogorov complexity metrics in assessing L2 proficiency: An information-theoretic approach. Frontiers in Psychology, 13, 1024147. https://doi.org/10.3389/fpsyg.2022.1024147
- Yoon, H. J. (2021). Challenging the connection between task perceptions and language use in L2 writing: Genre, cognitive task complexity, and linguistic complexity. *Journal of Second Language Writing*, 54, 100857. https://doi.org/10.1016/j.jslw.2021. 100857
- Yoon, H. J. (2018). Linguistic complexity in L2 writing revisited: Issues of topic, proficiency, and construct multidimensionality. *System*, *66*, 130–141. https://doi.org/10.1016/j.system.2017.03.007
- Zhang, X., & Lu, X. (2022). Revisiting the predictive power of traditional vs. fine-grained syntactic complexity indices for L2 writing quality: The case of two genres. *Assessing Writing*, *51*, 100597. https://doi.org/10.1016/j.asw.2021.100597

How to cite this article: Wang, H., Wang, G., Wang, N., & Wang, L. (2023). Effects of speaker types and L1 backgrounds on the linguistic complexity of learners' writing. *International Journal of Applied Linguistics*, 1–17. https://doi.org/10.1111/ijal.12526