

The Fine-Grained Syntactic Complexity in Research Abstracts: a Comparative Study of Vietnamese Students and Expert Native Writers across Disciplines

Bui Bien Thao Nhi, Nguyen Thi Quyen*



Use your smartphone to scan this QR code and download this article

ABSTRACT

Syntactic complexity, reflected in long sentences and multiple embedded clauses, is a key feature of high-quality academic writing as it allows writers to express complex ideas concisely and effectively. While much prior research has addressed syntactic complexity at a broad level, studies employing fine-grained measures remain limited. This study examines fine-grained syntactic complexity in research abstracts written by Vietnamese student writers, comparing them with abstracts authored by expert native English writers. Additionally, the study examines the differences between English majors and non-English majors. The analysis revealed clear differences between students and experts in both clausal and phrasal indices, alongside notable disciplinary variation. Collectively, the student writers demonstrated syntactic complexity approaching the level of expert authors, though English majors showed less advanced use of syntactic structures compared with non-English majors. The findings contribute to understanding disciplinary and language-background influences on academic writing and highlight implications for theory and pedagogy.

Key words: fine-grained syntactic complexity, disciplinary variation, language background, expert native writers, research abstracts

INTRODUCTION

In second language (L2) studies, complexity, accuracy, and fluency are often regarded as the three foundational dimensions used to assess learners' performance. Of these, complexity and accuracy represent learners' underlying grammatical and linguistic competence, while fluency reflects the extent to which that knowledge can be accessed and applied in real-time communication¹. Within this broader framework, syntactic complexity, understood as "the range and sophistication of grammatical resources displayed in language production" [2, p.82], has become a particularly salient indicator of progress and quality in L2 writing.

Researchers have operationalized the syntactic complexity through the two primary approaches: the large-grained and fine-grained measurement. Large-grained metrics, which include indices such as mean length of clause or T-unit, offer a broad view of structural elaboration. In contrast, fine-grained indices examine internal syntactic components, including clausal complements, noun phrase modifiers, and conjunction patterns, thereby providing a more detailed and sensitive account of writing sophistication. Contemporary studies increasingly favor fine-grained

measures for their ability to capture subtle distinctions in academic writing quality³⁻⁵.

Empirical evidence consistently links the higher syntactic complexity to greater writing proficiency. As demonstrated in³, lower-rated undergraduate essays contained markedly fewer complex clauses across multiple large-grained indicators. Disciplinary variation has also been reported: It has been found that research abstracts in Electronic Engineering were characterized by longer structural units and more sophisticated phrasal constructions than those in Education⁶. More broadly, the syntactic complexity has been found to vary with writing proficiency^{7,8}, disciplinary discourse practices^{9,10}, and language background^{11,12}.

For many Vietnamese undergraduates, writing a graduation thesis in English represents one of the most demanding academic milestones. The thesis requires not only mastery of disciplinary knowledge but also the ability to construct argumentation that meets international academic writing norms. Non-native writers reportedly frequently face challenges in achieving these norms, especially in managing syntactic and rhetorical conventions. Yet academic writing competence is increasingly crucial, not only for institutional assessment but also for gaining access to

International University, VNUHCM, Vietnam

Correspondence

Nguyen Thi Quyen, International University, VNUHCM, Vietnam

Email: ntquyen@hcmu.edu.vn

History

- Received: 14-10-2025
- Revised: 06-02-2026
- Accepted: 25-05-2026
- Published Online: 28-09-2026

DOI : <https://doi.org/10.32508/vnuhcmjssh.v10i3.1257>



Check for updates

Copyright

© VNUHCM Journal . This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Cite this article : Nhi B B T, Quyen N T. **The Fine-Grained Syntactic Complexity in Research Abstracts: a Comparative Study of Vietnamese Students and Expert Native Writers across Disciplines.** VNUHCMJ. Soc. Sci. Humanit. 2026;10(3):4127-4143.

the global scholarly community. Despite the growing international presence of L2 authors, linguistic expectations continue to hinder equitable participation in publication. Editorial guidelines and peer-review practices at leading journals often treat nativeness as an implicit measure of textual quality. Consequently, many manuscripts by non-native researchers are rejected for stylistic or linguistic reasons rather than deficiencies in research content¹³. Vietnamese doctoral students studying overseas, for instance, frequently report receiving corrective feedback focused mainly on language use¹⁴. Given this context, examining how Vietnamese EFL students construct academic texts in comparison with expert writers is both timely and necessary. In this study, expert native English writers are defined as individuals with advanced academic degrees obtained from universities in English-speaking countries and with publications in peer-reviewed outlets. By contrasting their abstracts with those written by Vietnamese students, the research aims to identify areas of convergence and divergence in the syntactic complexity, thereby informing pedagogical strategies for developing advanced academic writing proficiency.

Although numerous studies have investigated the syntactic complexity in relation to proficiency^{15,16}, disciplinary practice^{9,17}, and language background¹⁸⁻²⁰, empirical findings remain inconsistent. Additionally, few have employed the fine-grained indices capable of distinguishing subtle structural differences. Research into disciplinary variation is particularly limited, especially in Asian and EFL higher-education contexts.

To address these gaps, the present study analyzes the fine-grained syntactic complexity of research abstracts written by Vietnamese undergraduate students in the four academic fields, namely English Linguistics, Computer Science, Industrial Engineering and Systems, and Business Administration, and compares them with abstracts authored by expert native English writers. The study is guided by the two central research questions (RQ):

RQ1. How does the syntactic complexity differ between abstracts written by Vietnamese undergraduate students and those written by expert native English authors?

RQ2. In what ways does the syntactic complexity vary between Linguistics and non-Linguistics undergraduate abstracts?

This investigation contributes to a more comprehensive understanding of how linguistic background and disciplinary conventions interact in shaping the syntactic development among EFL academic writers. The

findings are expected to provide empirical benchmarks for Vietnamese learners, whose English exposure often relies primarily on expert-authored models rather than sustained communicative use. From a pedagogical perspective, the results highlight syntactic features that can be emphasized in academic writing instructions such as clausal complement structures, phrasal modifiers, and relative clause constructions to help students produce abstracts that align with international publication standards. In doing so, the study not only supports Vietnamese students in meeting institutional thesis requirements but also equips them with the syntactic awareness necessary for active participation in the global academic community.

LITERATURE REVIEW

The syntactic complexity is widely acknowledged as a core dimension in evaluating the L2 writing proficiency and development. It represents the degree to which writers elaborate and vary grammatical structures within their texts²¹ and is typically viewed as one component of the broader construct of linguistic complexity, which also encompasses lexical variation and sophistication²². Within this broader framework, the syntactic complexity plays a particularly critical role in determining the holistic quality of academic writing produced by L2 learners²³⁻²⁵. Researchers have examined the syntactic complexity across three analytical levels, namely global, clausal, and phrasal²³. These levels are commonly quantified through indices reflecting general structural elaboration, coordination, and subordination. Coordination-based measures tend to capture developmental features characteristic of beginner writing, whereas indices of subordination are more effective for identifying the syntactic growth at the intermediate stage. At higher proficiency levels, phrasal elaboration measures, which capture the density and internal sophistication of noun phrases, are considered more reliable indicators of the advanced academic writing ability.

Large-grained indices

Early research on the syntactic development in L2 writing has been largely dominated by large-grained indices, which assess broad structural dimensions such as production-unit length, the extent of subordination and coordination, and overall syntactic sophistication^{2,11,23}. Several studies have observed that writers producing longer clauses and sentences tend to achieve higher ratings for writing quality,

whereas frequent use of short, simple sentence patterns is often associated with less advanced compositions^{7,26}. This assumption that greater length corresponds to greater syntactic complexity has guided early studies and has made large-grained measures a popular choice in evaluating student writing performance^{3,6,27}. However, these metrics have been questioned for their interpretive limitations. Because they rely primarily on structural length (e.g., T-unit or sentence length), they may obscure the underlying grammatical mechanisms that truly contribute to complexity. As noted in^{23,25,28}, such length-based indices provide only a surface-level estimate of sophistication and fail to reveal the functional variety or intricacy of syntactic constructions within the text.

Fine-grained indices

In response to the limitations of traditional, large-scale syntactic complexity measures, recent research has turned toward fine-grained indices that allow for more accurate representation of clausal and phrasal structures. Rather than aggregating multiple constructions into a single composite value, these indices quantify each syntactic structure separately, thereby addressing the issue of over-generalization and preserving the distinct sources of complexity within texts²⁹. Fine-grained clausal indices are typically operationalized through the frequency of different clause dependency types, providing a detailed account of subordination²⁵. Although phrasal indices are often stronger predictors of academic writing quality, clausal measures contribute complementary explanatory value and have also been shown to correlate significantly with proficiency^{25,30}. Fine-grained phrasal indices, in turn, are commonly assessed along the three dimensions: the frequency of dependent items, the mean number of dependents per noun phrase, and the internal distribution of modifiers within noun phrases^{31,32}. The higher mean values in these phrasal measures are associated with more advanced writing²⁵.

Building on this clausal–phrasal distinction, a recurring conclusion in research on syntactic development is that the advanced academic writing is more reliably indexed by phrasal elaboration than by sentence length or clausal subordination. Developmental syntheses and corpus-based studies suggest that, as proficiency and academic experience increase, writers tend to rely less on extensive clausal embedding and more on compressed noun-phrase structures that enable dense and precise information packaging^{2,23,30,33}. In fine-grained terms, advanced academic prose is therefore often associated with higher frequencies and/or

greater density of noun phrase modification, including (i) adjectival modification, (ii) prepositional and nominal modification, and (iii) relative-clause modification, as well as broader indicators capturing the number and distribution of dependents within nominals^{25,31,34}. Because these constructions function as key mechanisms of informational compression in academic registers, fine-grained phrasal indices are widely treated as central indicators of advanced syntactic sophistication, while fine-grained clausal indices (e.g., complements, adverbial clauses/modifiers) are typically interpreted as complementary evidence reflecting different elaboration strategies conditioned by genre and discipline^{25,30,33}.

Despite their analytical precision, the fine-grained approaches were historically less common than the large-grained measures, largely due to methodological constraints. Earlier studies lacked the computational tools capable of reliably identifying and quantifying detailed syntactic features, making manual annotation labor-intensive and difficult to replicate across corpora^{3,30,34}. In contrast, the broader indices such as the mean length of T-unit or clauses per sentence were easier to compute and interpret, contributing to their widespread adoption in early L2 writing research³⁵. With advances in computational linguistics, however, the fine-grained analysis has become increasingly feasible. Tools, including the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC) enable the consistent automated extraction of fine-grained indices from large corpora²⁵. Consequently, the fine-grained measures are now more widely used in studies of academic writing^{19,20,36}. For instance,³⁷ showed that higher-quality IELTS Task 2 essays exhibited both greater clausal complexity and richer noun phrase structure.

Multiple factors influencing syntactic complexity in academic writing

An expanding body of scholarship has explored the variables shaping the syntactic complexity in academic writing, with the particular attention to writing experience, disciplinary context, and language background. These three dimensions also constitute the focus of the present research. The study contrasts the writing of expert native English authors who possess considerable experience in producing research articles and navigating academic discourse with that of Vietnamese undergraduates, who are comparatively new to academic writing conventions. Within this student cohort, comparisons are further drawn between English Language majors and students from

non-linguistic disciplines. Because standardized proficiency data were unavailable, writing proficiency itself was not treated as a separate analytical variable. Syntactic maturity has consistently been associated with increased writing experience. It has been noted in³⁸ that university students often struggle to construct grammatically complex discourse, and that phrasal elaboration tends to expand as writers gain more exposure to academic tasks. This development is typically attributed to the communicative pressure of condensing dense information within concise textual limits, a hallmark of academic genres^{34,38,39}. Numerous empirical investigations have confirmed that more experienced writers are likely to achieve higher levels of academic writing sophistication^{8,15,16,40}. Specifically, expert authors frequently employed advanced phrasal constructions, whereas novice writers relied more on simpler forms^{16,38,40}. Nonetheless, the results remain uneven. As shown in¹⁵, Thai EFL writers made the heavier use of noun pre-modifiers and distinct patterns of preposition usage, while⁴⁰ detected the minimal variation in prepositional structures but observed a stronger dependence on attributive adjectives among experienced writers.

The influence of language background on the syntactic complexity remains a subject of debate. Some studies, such as³⁶, reported no significant differences between advanced L2 writers and native English speakers, as observed among Chinese doctoral students. In contrast, other investigations revealed the substantial variation across both clausal and phrasal domains^{11,18–20,41}. Even within a single dimension, namely subordination, findings diverge:¹¹ found that EFL learners employed fewer subordinate clauses than native writers, whereas¹⁸ reported the opposite. Similar inconsistencies emerge for phrasal elaboration, some studies attribute the higher phrasal density to native writers^{18–20}, while others identify this trend in L2 texts³⁹.

Genre also exerts a strong influence on syntactic patterning. Previous studies have primarily examined published journal articles^{19,20,36} or argumentative essays^{11,18}. By contrast, undergraduate theses, typically written by novice scholars, remain comparatively neglected in syntactic complexity research. Moreover, many studies have analyzed entire research articles^{8,20}, which may obscure syntactic distinctions among specific rhetorical sections⁹. To address these gaps, the present study isolates research abstracts as the analytical focus. Abstracts provide a concise yet information-dense site for syntactic analysis due to

their standardized rhetorical structure and communicative precision.

Disciplinary differences also shape syntactic preferences. Academic fields are known to influence the surface structure of texts⁴². Using large-grained indices,⁹ showed that Applied Linguistics writing displayed the greater syntactic complexity than other fields, while⁴³ confirmed similar tendencies with phrasal indices. However, other findings are less consistent:¹⁷ reported no significant differences between pure and applied sciences.¹⁹ found that Applied Linguistics and Economics favored clausal complexity, while Chemistry relied more on phrasal structures. Despite these insights, many of these studies relied on the traditional large-grained measures, which have been criticized for interpretive limitations³⁰. Notably, few have explicitly compared Linguistics majors, who might be expected to leverage the metalinguistic awareness, with students from other fields. Investigating this distinction offers the potential to clarify whether disciplinary knowledge translates into more advanced syntactic performance.

Research gap

Despite the growing literature, several gaps remain. First, the fine-grained indices, though theoretically powerful, are still less frequently applied than the large-grained measures, particularly in EFL contexts. Second, the disciplinary variation in L2 student writing has not been thoroughly examined, especially when comparing Linguistics majors, who might be assumed to have the greater metalinguistic awareness, with students from non-linguistic disciplines. Third, many studies have analyzed full-length articles or essays, while undergraduate thesis abstracts, texts that demand conciseness and linguistic precision, remain underexplored.

To address these limitations, the present study explores the fine-grained syntactic complexity in Vietnamese undergraduate thesis abstracts, comparing them with those produced by expert native English writers. It further examines disciplinary differences between Linguistics and non-Linguistics student writing. By centering on the abstract, a section that demands both concision and complexity, the study offers a focused perspective on how language background, academic experience, and disciplinary orientation interact to shape the syntactic sophistication in academic writing.

METHOD

This study adopts a quantitative corpus-based comparative design to examine the fine-grained syntac-

tic complexity in research abstracts produced by Vietnamese undergraduate EFL writers and expert native English authors across disciplines. The syntactic complexity is operationalized using the fine-grained clausal and phrasal indices, allowing the systematic comparison of syntactic patterning across writer groups.

Corpus Description

The corpus for this study consisted of 210 research abstracts drawn from the four academic disciplines: Business Administration (BA); Computer Science (CS), Industrial Engineering and Systems (IE), and English Linguistics (EL). These texts were produced by the two distinct writer groups: (1) Vietnamese undergraduates enrolled at a Vietnamese university, and (2) expert native English (L1) writers. A purposive and balanced sampling strategy was adopted to ensure equal representation across writer groups and disciplines, with random selection applied at the final stage when sub-samples were required for disciplinary comparisons. All texts were selected in accordance with the clearly defined inclusion and exclusion criteria (see Table 1), adapted primarily from ^{16,19,20}, with the minor modifications to suit the present research context and objectives.

To best capture the current level of academic writing proficiency among Vietnamese students, the dataset incorporated the most recently available abstracts, irrespective of their year of completion (Group 1). Each student abstract underwent screening for the potential AI-generated content using Turnitin, and only those with the less than 20% detected AI text were retained. Although the accuracy of AI-detection systems remains a matter of scholarly debate ⁴⁴, Turnitin continues to be recognized as one of the more established and reliable tools for identifying both plagiarism and AI-assisted writing ^{45,46}. For the expert L1 corpus (Group 2), only abstracts published in 2022 or earlier were included, predating the public release of generative AI technologies such as ChatGPT (launched November 30, 2022). This precaution ensured that all selected texts were human-authored and reduced the dependence on automated authenticity checks, whose reliability remains contested ⁴⁴.

For Research Question 1, which explores the differences in the syntactic complexity between expert native English writers and Vietnamese undergraduate writers, particular care was taken to ensure the accuracy and validity of author identification. Unlike previous studies that inferred language background solely from an author's English name or institutional

affiliation, this research applied more rigorous selection standards. Specifically, expert writers were required to have completed all three academic degrees, i.e., Bachelor's, Master's, and Doctoral, at institutions located in English-as-a-native-language (ENL) countries. This condition assumes sustained academic immersion in English, encompassing advanced exposure to complex academic genres and discourse practices. As emphasized in ⁴⁷, the mastery of academic writing extends beyond mere linguistic proficiency to include genre awareness, disciplinary understanding, and rhetorical control, all of which develop through long-term academic engagement. Accordingly, expert abstracts were drawn from Scimago-ranked Q1–Q2 journals to represent benchmark exemplars of discipline-appropriate abstract writing under high-stakes publication conditions. Importantly, the study is not designed to match the two writer groups by proficiency or cognition; rather, it adopts a benchmark-comparison design commonly used in corpus-based academic writing research, in which expert writing serves as a reference model. This design allows the analysis to identify which fine-grained syntactic resources are the characteristic of expert abstract writing and to assess the extent to which advanced EFL undergraduate abstracts approximate these expert patterns within the same academic genre.

For Research Question 2, which investigates the disciplinary differences in syntactic complexity between English Linguistics (EL) and non-Linguistics disciplines, it was necessary to confirm that each abstract was properly aligned with its respective academic domain. The non-EL group included Computer Science (CS), Industrial Engineering and Systems (IE), and Business Administration (BA), fields selected based on the availability of sufficient, recent student abstracts archived in the university's official digital repository. The sample sizes (25 abstracts per writer group in CS, IE, and BA, and 30 per group in EL) were determined by the availability of eligible texts meeting the strict inclusion criteria, the need for balanced group comparisons, and the consistency with sample sizes adopted in prior corpus-based studies of the fine-grained syntactic complexity.

As described in Table 1, all abstracts were collected from the two primary sources: the university's official KEEP repository (student theses) and Scimago-indexed journals (expert L1 authors). Accordingly, expert abstracts were selected exclusively from Scimago-ranked Q1 and Q2 journals, representing the top quartiles in research quality; journals in these categories are characterized by stringent peer-review standards and high linguistic preci-

Table 1: The Inclusion and Exclusion Criteria [Source: authors]

Inclusion criteria	Exclusion criteria
The abstract must be written in English. The abstract must correspond to one of the following fields of study: English Linguistics (EL), Computer Science (CS), Industrial Engineering and Systems (IE), or Business Administration (BA). Each abstract must contain a minimum of 150 words. • The author background: – Group 1: Vietnamese undergraduate students enrolled at the selected university. – Group 2: Writers with English-sounding names affiliated with ENL institutions who obtained all academic degrees (BA/BSc, MA, PhD) from ENL universities. • The source of data: – Group 1: Abstracts published on the KEEP repository, the university’s official digital archive of graduation theses; for English Linguistics, abstracts may also be retrieved from the School of Language’s website. – Group 2: Abstracts published in high-ranking (Q1–Q2) journals indexed by Scimago. • Publication timeframe: – Group 1: Abstracts submitted between 2020 and 2025. – Group 2: Abstracts published between 2017 and 2022.	Texts that do not clearly belong to one of the four target disciplines (e.g., cross-disciplinary work combining IE and BA). Abstracts exceeding 300 words in length. Papers categorized as theoretical reviews or non-empirical research. • The author background: – Group 1: Students who are not native Vietnamese (e.g., biracial or international students). – Group 2: English-named writers not affiliated with ENL institutions or whose educational history cannot be verified.

Table 2: The Description of the Corpus [Source: authors]

	CS		IE		BA		EL		Total
	L1	EFL	L1	EFL	L1	EFL	L1	EFL	
No. of texts	25	25	25	25	25	25	30	30	210
No. of words	5226	5814	5113	5250	4861	5626	5811	6318	44019
Mean words	209.04	232.56	204.52	210	194.44	225.04	193.7	210.6	209.61

sion, making them the appropriate exemplars of advanced, discipline-appropriate academic writing^{33,48}. All texts were screened against predefined inclusion and exclusion criteria, and student abstracts were additionally checked for potential AI-generated content using Turnitin, with only texts below the specified threshold retained. A summary of the descriptive statistics for the entire corpus is presented in Table 2.

Data Analysis

All collected abstracts were compiled in a master spreadsheet containing metadata, including author names, journal titles, word counts, and references to facilitate traceability and quality control. Each abstract was checked for typographical and formatting errors, and extraneous elements (e.g., parenthetical citations, ordinal numbers, and bracketed abbreviations) were removed. Cleaned texts were then converted to plain-text format as required for computational processing.

The Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC)

The analysis of syntactic features in this study was conducted using the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC), originally developed by³¹. This computational tool employs the Stanford neural-network dependency parser to identify syntactic relations with an accuracy rate of roughly 90%, allowing for reliable large-scale syntactic parsing. TAASSC was selected over other available programs, such as Biber Tagger and Coh-Metrix, due to its open accessibility, efficient processing speed, and comprehensive capacity to extract a wide array of the fine-grained syntactic indices. The system computes both clausal and phrasal complexity measures, which are grounded in the well-established frameworks of syntactic sophistication in L2 writing^{23,34}. Distinguishing itself from the traditional large-grained methods, TAASSC (i) measures syntactic length by calculating dependents-per-clause and dependents-per-nominal ratios instead of raw

word counts, (ii) analyzes individual syntactic structure types separately rather than merging them into aggregated categories, and (iii) defines clauses as a main verb with its associated dependents, providing a more accurate linguistic representation of syntactic structure²⁵.

The Measures of Syntactic Complexity

Following the framework established by⁴⁹ and incorporating the modifications suggested by¹⁹, this research examined a total of 21 fine-grained syntactic indices, including nine clausal and twelve phrasal measures, which were selected for their representativeness of advanced academic prose. These indices were computed using TAASSC, enabling the detailed measurement of both the range and density of syntactic elaboration. The clausal indices (see Table 3) quantify the frequency of dependent constructions, encompassing finite and non-finite clauses³¹. Indices such as dependents per clause, adverbial clauses per clause, and clausal complements per clause were employed to capture variation in subordination and coordination patterns. Finite complements were categorized under the clausal complement index, whereas non-finite complements were represented through the open clausal complement index. Clauses functioning as noun modifiers in non-finite form contributed to both the dependents-per-clause and dependents-per-nominal indices. Moreover, adverbial and prepositional phrases serving adverbial functions were incorporated via the adverbial modifier index. Collectively, these features provide a comprehensive representation of clausal complexity, offering insights that phrasal-level measures alone cannot capture¹⁹.

The phrasal indices (see Table 4) evaluate the internal complexity of noun phrases by measuring both the density and diversity of dependents, comprising determiners, adjectival modifiers, prepositional and relative clauses, possessives, as well as verbal, nominal, and adverbial modifiers²⁵. The initial group of indices reflects the average number of dependents across noun phrases and their associated grammatical functions, including nominal subjects, prepositional objects, and nominal complements. The remaining indices capture the frequency of specific dependent types, e.g., adjectival or relative clause modifiers, within noun phrases^{19,31}. Consistent with⁵⁰, pronouns were omitted from the analysis due to their limited occurrence in academic writing. Typically, higher mean scores on these measures denote greater informational density and structural sophistication, hallmarks of expert-level academic prose.

The Statistical analysis

TAASSC outputs were aggregated to obtain mean values for each syntactic index. Descriptive statistics were used to summarize clausal and phrasal complexity across writer groups and disciplines. To ensure comparability, sample sizes were balanced: for RQ1, 105 abstracts by native English authors and 105 by Vietnamese students were analyzed; for RQ2, 30 abstracts each from English Linguistics and non-Linguistics disciplines (CS, IE, BA) were randomly selected. All statistical analyses were conducted using STATA. Because the Shapiro-Wilk tests indicated that most syntactic indices deviated from normality and several measures displayed skewed distributions with outliers, a non-parametric approach was adopted. The Mann-Whitney U test was therefore used, as it compares group differences based on rank distributions rather than means and does not assume normality. Although Welch's *t*-test is robust to unequal variances, it remains mean-based and sensitive to distributional irregularities; thus, Mann-Whitney U was considered more appropriate for the present data. Statistical significance was established at a threshold of $p < .05$.

RESULTS

The Descriptive Statistics of Syntactic Complexity

From the total corpus of 210 research abstracts (105 by expert L1 writers and 105 by EFL learners), the descriptive statistics for the 21 fine-grained syntactic indices are presented in Figure 1. Overall, most indices show the minimal variation between groups, with only six indices displaying notably higher mean values in the expert group. The standard deviations for all measures were below 2, indicating stable distributions, with the largest (1.9) observed in *av_ncomp_deps_NN*.

For the disciplinary comparison, 10 abstracts from each non-EL discipline were randomly selected. Figure 2 displays the descriptive statistics for the EL and non-EL groups. As with the cross-language comparison, the standard deviations were consistently below 2. The EL group showed noticeably higher means for four indices, although the global distribution patterns across both groups were largely similar.

The Differences in Clausal Complexity

Table 5 presents the results of the Mann-Whitney U test comparing the clausal complexity between expert L1 writers and EFL writers. Native experts consistently demonstrated higher mean ranks

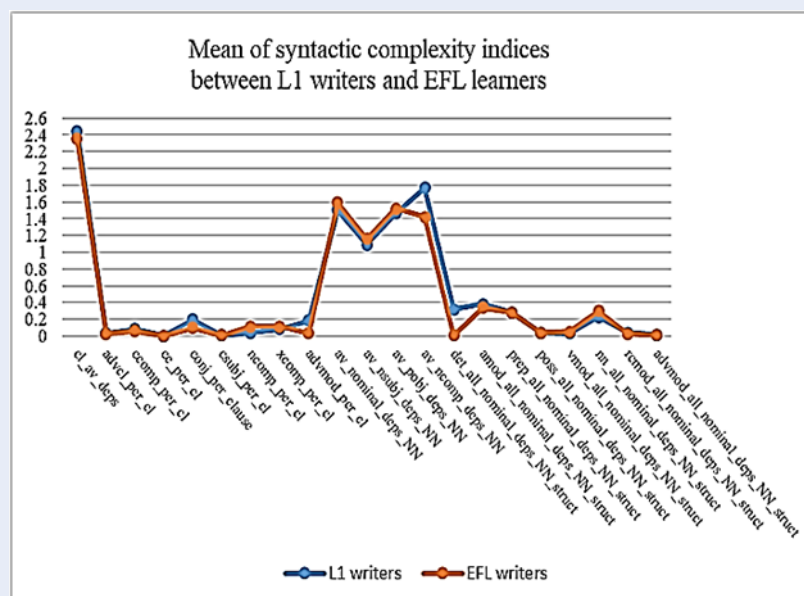


Figure 1: The Differences between Expert L1 Writers and EFL Writers [Source: authors]

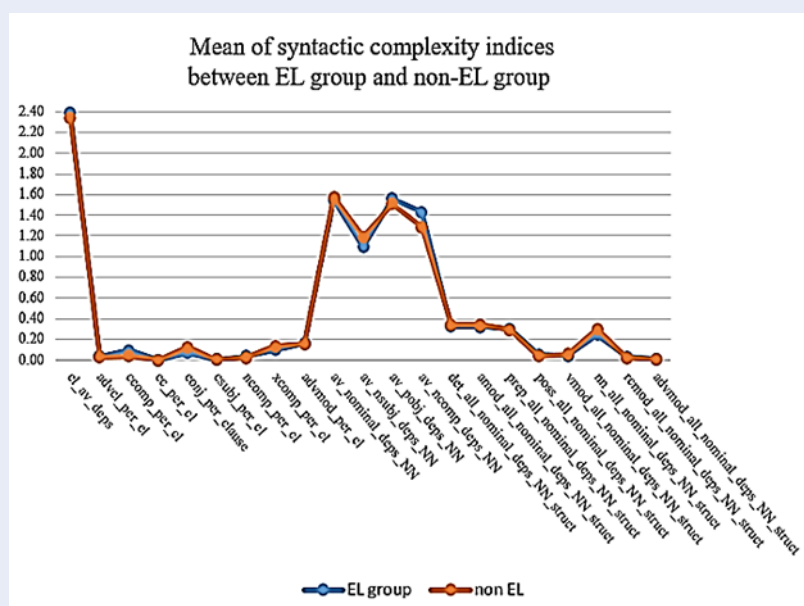


Figure 2: The Differences between the EL Group and the Non-EL Group [Source: authors]

Table 3: The Clausal Complexity Indices [Source: adapted from³¹

No.	Index name	In-text name
1	cl_av_deps	dependents per clause
2	conj_per_clause	conjunctions per clause
3	csubj_per_cl	clausal subjects per clause
4	advcl_per_cl	adverbial clauses per clause
5	advmod_per_cl	adverbial modifiers per clause
6	ccomp_per_cl	clausal complements per clause
7	ncomp_per_cl	nominal complements per clause
8	xcomp_per_cl	open clausal complements per clause
9	cc_per_cl	clausal coordinating conjunctions per clause

Table 4: The Phrasal Complexity Indices [Source: adapted from³¹

No	Index name	In-text name
1	av_nominal_deps_NN	nominal dependents per nominal (no pronouns)
2	av_nsubj_deps_NN	dependents per nominal subject (no pronouns)
3	av_pobj_deps_NN	dependents per object of the preposition (no pronouns)
4	av_ncomp_deps_NN	dependents per nominal complement (no pronouns)
5	det_all_nominal_deps_NN_struct	determiners per nominal (no pronouns)
6	amod_all_nominal_deps_NN_struct	adjectival modifiers per nominal (no pronouns)
7	prep_all_nominal_deps_NN_struct	prepositions per nominal (no pronouns)
8	poss_all_nominal_deps_NN_struct	possessives per nominal (no pronouns)
9	vmod_all_nominal_deps_NN_struct	verbal modifiers per nominal (no pronouns)
10	nn_all_nominal_deps_NN_struct	nouns as a nominal dependent per nominal (no pronouns)
11	rcmod_all_nominal_deps_NN_struct	relative clause modifiers per nominal (no pronouns)
12	advmod_all_nominal_deps_NN_struct	(non-clausal) adverbial modifiers per nominal (no pronouns)

across nearly all indices, with the exception of xcomp_per_cl. The statistically significant differences ($p < .05$) emerged in the three measures, namely dependents per clause, open clausal complements per clause, and adverbial modifiers per clause. These findings indicate that expert authors tended to use a greater number of dependent and adverbial clauses, while EFL writers showed a comparatively higher frequency of open clausal complements.

Table 6 shows the disciplinary comparison between EL and non-EL student groups. The EL group displayed higher mean ranks across four indices; however, only three measures differed significantly: clausal complements per clause (higher in EL), conjunctions per clause, and open clausal complements per clause (both higher in non-EL).

The Differences in Phrasal Complexity

Table 7 illustrates that five out of the twelve phrasal indices showed statistically significant differences between expert L1 writers and EFL writers. The expert group exhibited greater use of adjectival modifiers and relative clause modifiers, whereas the EFL group produced a higher frequency of nominal dependents, verbal modifiers, and nouns functioning as nominal dependents. These patterns suggest that native English writers tend to favor adjectival and relative-clause-based modification, while EFL writers more frequently construct embedded nominal or verbal structures to achieve syntactic elaboration.

Finally, Table 8 presents the phrasal complexity results for EL and non-EL groups. Both groups exhibited comparable mean ranks across most measures; how-

Table 5: The Results of Clausal Complexity Measures Differences between Expert L1 Writers and EFL Writers [Source: authors]

Clausal indices	L1 vs EFL (N = 210)				
	Mean rank		U	Z	p-value
	L1	EFL			
cl_av_deps	115.057	95.94	4509	2.279	.022*
advcl_per_cl	110.295	100.705	5009	1.211	.226
ccomp_per_cl	113.071	97.929	4717.5	1.820	.069
cc_per_cl	105.543	105.457	5508	0.035	.715
conj_per_clause	109.014	101.986	5143.5	0.839	.403
csubj_per_cl	107.257	103.743	5328	0.679	.475
ncomp_per_cl	110.895	100.105	4946	1.353	.176
xcomp_per_cl	87.224	123.776	7431.5	-4.363	.000*
advmod_per_cl	115.814	95.186	4429.5	2.460	.014*

Note. * $p < .05$ indicates statistical significance

Table 6: The Results of Clausal Complexity Measures between EL Group & Non-EL Group [Source: authors]

Clausal indices	EL group and non-EL group (N = 60)				
	Mean rank		U	Z	p-value
	EL	non-EL			
cl_av_deps	32.183	28.817	399.5	0.747	.46
advcl_per_cl	32.267	28.733	397	0.818	.418
ccomp_per_cl	37.267	23.733	247	3.031	.0021*
cc_per_cl	30.5	30.5	450	0	1
conj_per_clause	24.267	36.733	637	-2.768	.0051*
csubj_per_cl	30	31	465	-0.357	.761
ncomp_per_cl	32.35	28.650	394.5	0.891	.378
xcomp_per_cl	25.717	35.283	593.5	-2.123	.034*
advmod_per_cl	29.5	31.5	480	-0.444	.662

Note.* $p < .05$ indicates statistical significance

ever, significant differences were observed in two indices. The non-EL group showed higher frequencies of nouns as nominal dependents, while the EL group demonstrated the greater use of adverbial modifiers per nominal.

DISCUSSION

The Differences between Expert L1 Writers and Vietnamese Student Writers

The results revealed that expert L1 writers generally used clausal structures more frequently across most indices, except for open clausal complements per

clause. This finding diverges from³⁸, who reported that higher academic levels are often associated with reduced clausal subordination, but converges with⁵¹, who found that expert texts display more indicators of clausal elaboration than learner writing.

For the phrasal complexity, student writers demonstrated higher average values in eight of the twelve indices, three of which showed statistically significant differences. This pattern suggests that the Vietnamese students' phrasal complexity approximates the expert-level usage. However, L1 writers surpassed students in the use of adjectival and relative clause modifiers, indicating that these features re-

Table 7: The Results of Phrasal Complexity Measures Differences between Expert L1 & EFL Writers [Source: authors]

Phrasal indices	L1 vs EFL (N = 210)				
	Mean rank		U	Z	p-value
	L1	EFL			
av_nominal_deps_NN	90.42	120.58	7095.5	-3.6	.0003*
av_nsubj_deps_NN	97.51	113.49	6351.5	-1.91	.056
av_pobj_deps_NN	99.14	111.86	6180	-1.52	.13
av_ncomp_deps_NN	110.54	100.45	4983	1.3	.195
det_all_nominal_deps_NN_struct	98.96	112.04	6199.5	-1.56	.119
amod_all_nominal_deps_NN_struct	117.75	93.25	4226	2.92	.003*
prep_all_nominal_deps_NN_struct	103.9	107.1	5680.5	-0.38	.704
poss_all_nominal_deps_NN_struct	102.67	108.33	5809.5	-0.68	.5
vmod_all_nominal_deps_NN_struct	93.63	117.37	6759	-2.83	.005*
nn_all_nominal_deps_NN_struct	85.77	125.23	7584	-4.71	.000*
rcmod_all_nominal_deps_NN_struct	120.59	90.41	3928	3.61	.0003*
advmod_all_nominal_deps_NN_struct	111.34	99.66	4899	1.57	.116

Note.* p < .05 indicates statistical significance

Table 8: The Results of Phrasal Complexity Measures Differences between EL Group & Non-EL Group [Source: authors]

Phrasal indices	EL vs non-EL (N = 60)		U	Z	p-value
	Mean rank				
	EL	non-EL			
av_nominal_deps_NN	27.95	33.05	526.5	-1.13	.258
av_nsubj_deps_NN	28.55	32.78	518.5	-1.01	.315
av_pobj_deps_NN	31.75	29.25	412.5	0.55	.584
av_ncomp_deps_NN	31.18	29.82	429.5	0.34	.739
det_all_nominal_deps_NN_struct	28.7	32.26	503	-0.8	.43
amod_all_nominal_deps_NN_struct	29.22	31.78	488.5	-0.57	.574
prep_all_nominal_deps_NN_struct	32.17	28.83	400	0.74	.465
poss_all_nominal_deps_NN_struct	33.4	27.6	363	1.29	.2
vmod_all_nominal_deps_NN_struct	26.48	34.51	570.5	-1.78	.075
nn_all_nominal_deps_NN_struct	25.12	35.88	611.5	-2.39	.016*
rcmod_all_nominal_deps_NN_struct	34.4	26.6	333	1.74	.083
advmod_all_nominal_deps_NN_struct	35.05	25.95	313.5	2.38	.019*

Note.* p < .05 indicates statistical significance

main key areas for pedagogical development. Given that³⁰ identified phrasal elaboration as a hallmark of advanced writing, these results imply that the students' abstracts already reflect a relatively high level of syntactic maturity. Moreover, the observed pattern, i.e., decreased clausal subordination alongside increased phrasal complexity, corresponds with the developmental trajectory proposed for advanced L2 writers^{2,23,28}.

As noted by³⁰, the advanced academic writing is typically distinguished by a frequent and skillful use of phrasal complexity. In light of this, the abstracts produced by the student writers can be interpreted as demonstrating a relatively high degree of syntactic sophistication. Furthermore, the combined findings on clausal and phrasal complexity suggest that these students follow the trajectory of the developmental progression commonly observed among advanced L2/EFL writers, one marked by a decline in clausal subordination and a greater dependence on phrasal elaboration^{2,23,28}.

These findings also challenge earlier assumptions that non-native writers inherently struggle with academic composition¹³. The students' sophisticated use of phrasal constructions may be attributed to Vietnam's early English education policies and to the English-Medium Instruction (EMI) environment of their university, where continuous exposure to English across academic and disciplinary contexts facilitates syntactic development.

The analysis of dependent types contributes to the debate between¹⁵ and⁴⁰. Consistent with¹⁵, expert writers used adjectival and relative clause modifiers more frequently, while students relied more on verbal modifiers and nouns as nominal dependents. The latter corroborates findings in¹⁵ but contradicts⁴⁰, which associated noun pre-modification with expert writing. Furthermore, the students' lower use of relative clauses corresponds with the observations of EFL learners' syntactic preferences¹⁵.

As far as cross-language differences are concerned, in clausal complexity, L1 writers employed more dependent clauses overall, supporting findings reported in³⁶ that native speakers use greater subordination than EFL writers, even at advanced levels. The L1 group also used more adverbial clauses, whereas EFL writers favored open clausal complements, a pattern opposite to results in¹⁹, which found L1 writers relying more on open complements. Although other indices showed slight differences favoring L1 writers, these were not statistically significant, indicating only a modest advantage in clausal elaboration, corroborating the findings of¹¹.

In phrasal complexity, students' higher use of nominal dependents contradicts²⁰ but supports³⁹. While L1 writers characteristically employ a wider variety of noun phrase structures, both groups exhibited diverse modification patterns across indices, suggesting that the fine-grained measures capture subtler stylistic variation than previously noted. The frequent use of nouns as nominal dependents among student writers echoes³⁸, who found similar tendencies among Arabic and Chinese writers. The collective phrasal findings partly corroborate¹⁹, particularly regarding dependents per nominal and the distribution of adjectival and relative modifiers.

These results diverge from¹⁸, who found that L1 texts displayed greater phrasal elaboration while L2 texts favored subordination. The discrepancy may stem from methodological differences: the present study used fine-grained indices, while earlier research relied on large-grained measures that obscure distinctions between clausal and phrasal sources of complexity¹⁹.

The Differences between EL and Non-EL Groups

Disciplinary variation revealed roughly balanced differences in both clausal and phrasal measures. Although some indices reached statistical significance, the global syntactic difficulty was comparable between EL and non-EL writers.

For clausal complexity, the three significant distinctions emerged: the non-EL group used more conjunctions and open clausal complements, whereas the EL group produced more clausal complements. According to³⁰, open clausal complements are characteristic of advanced academic prose, while finite complements are more typical of conversational or simpler registers⁵². These results partly align with¹⁹ and⁴³ regarding complement usage, though they contrast with their broader claims about disciplinary trends.

Contrary to expectations, the EL group did not outperform non-EL students in phrasal complexity. Only two indices differed significantly: the EL group used fewer nouns as nominal dependents but more adverbial modifiers per nominal. The former structure is typical of academic registers⁵³, whereas the latter is more common in lower-proficiency writing⁵⁴. These patterns reinforce¹⁹, suggesting that linguistic knowledge alone does not guarantee syntactically advanced writing performance.

It may be implicated from the findings that specialized linguistic training may not translate directly into syntactic sophistication in academic abstracts. This contrasts with⁹ and⁴³, who reported that Applied

Linguistics writers employed more complex structures than those in other disciplines. Methodological differences, particularly the use of fine-grained rather than large-grained indices, may account for these contrasting outcomes. Importantly, these results should be interpreted within the specific institutional context of the research. The data were collected at a multi-disciplinary university, where the structure and intensity of language-major curricula may differ from those at institutions specializing in language and linguistics training. Consequently, the observed syntactic patterns among English Linguistics students are not intended to be generalized to EL programs across Vietnam, and variation may be expected across institutions with differing curricular emphases and student proficiency profiles.

From a pedagogical perspective, the similarities and differences observed between the present findings and prior studies offer concrete insights into Vietnamese EFL learners' syntactic development in academic writing. The results suggest that while advanced learners can achieve considerable phrasal density, targeted support remains necessary for developing discipline-appropriate modifier types and balanced clausal-phrasal elaboration. For Vietnamese teachers of English, particularly in academic writing and thesis-preparation courses, these findings support instructional approaches that move beyond sentence-length metrics to focus on specific fine-grained syntactic features. For example, students can be asked to revise draft abstracts by diagnosing noun phrase structures (e.g., identifying over-reliance on nominal stacking or verbal modifiers) and restructuring them using alternative modifiers such as adjectival or relative clause modification, guided by expert-model abstracts. In this way, the fine-grained clausal and phrasal indices can be used diagnostically to support genre-based instruction and to help learners calibrate their syntactic choices to conventions of expert academic writing.

CONCLUSION

This study examined the fine-grained syntactic complexity in research abstracts written by expert native English authors and Vietnamese undergraduate EFL writers, alongside comparisons between English Linguistics and non-Linguistics student groups. Using a corpus of 210 abstracts and analyzing data through the TAASSC, the research assessed the 21 syntactic indices, comprising nine clausal and twelve phrasal measures, based on established frameworks^{25,30,31}. The findings revealed both commonalities and distinctions across linguistic backgrounds and disciplinary domains. Native English writers achieved

higher mean scores on most clausal indices, particularly dependents per clause, adverbial clauses per clause, and adverbial modifiers per clause, demonstrating greater command of subordination and dependency structures. In contrast, Vietnamese students made more frequent use of open clausal complements, verbal modifiers, and nouns functioning as nominal dependents, leading to denser phrasal constructions in aggregate. This pattern suggests that native writers tend to expand meaning through clausal embedding, whereas EFL writers achieve the syntactic elaboration primarily via phrasal compression.

Regarding disciplinary variation, EL and non-EL students displayed broadly similar levels of syntactic development, with statistically significant differences in only a few features. The EL group employed more clausal complements and adverbial modifiers, while non-EL students favored open clausal complements, conjunctions, and nouns as nominal dependents, structures typically linked to more advanced written academic discourse⁵³. Collectively, these findings suggest that the syntactic sophistication in undergraduate writing reflects not only linguistic knowledge and training, but also the influence of disciplinary writing conventions and genre-specific exposure.

The findings contribute to ongoing discussions on syntactic development in academic writing, particularly in EFL contexts. By integrating the fine-grained syntactic measures, the study offers a more nuanced perspective than earlier research relying on the large-grained indices such as mean length of T-unit or clause. This methodological advancement helps clarify inconsistencies in prior studies^{11,19} and demonstrates that syntactic sophistication cannot be fully captured through surface-level measures of length or subordination. Moreover, the observation that EFL writers sometimes surpass expert L1 writers in specific phrasal indices challenges the long-standing assumption that nativeness equates to syntactic superiority. The results lend support to a non-linear model of syntactic development, in which learner progress through varying balances of clausal and phrasal elaboration rather than following a single universal trajectory³⁰. These findings suggest that syntactic advancement should be viewed as multidimensional, sensitive to linguistic environment, academic exposure, and genre-specific expectations rather than solely to language background. Finally, the disciplinary differences observed in this study highlight the influence of academic field and rhetorical context on syntactic choices. While Linguistics students might be

expected to demonstrate superior syntactic awareness, the comparable or even stronger performance of non-EL students indicates that theoretical linguistic knowledge does not automatically translate into more sophisticated writing performance. This resonates with³³, who noted that disciplinary demands and genre conventions play a central role in shaping syntactic complexity.

From a pedagogical standpoint, the findings suggest several implications for EFL academic writing instruction. First, the demonstrated ability of Vietnamese undergraduate students to produce advanced phrasal structures highlights the potential effectiveness of early English instruction and the benefits of studying in an English-Medium Instruction (EMI) environment. Continuous exposure to authentic academic input and writing practices appears to enhance students' ability to construct syntactically dense and information-rich sentences. However, their relative underuse of clausal subordination and certain modifier types indicates areas where explicit instruction remains necessary. Writing pedagogy in both Linguistics and non-Linguistics programs should therefore move beyond abstract grammar teaching to emphasize genre-based, syntax-aware instruction, enabling students to apply linguistic principles within authentic academic contexts. As suggested by⁵⁵, incorporating targeted syntactic exercises into writing practice can significantly improve learners' writing development. Foundational writing courses may integrate research and proposal writing tasks that explicitly model and reinforce advanced syntactic constructions. Subsequently, discipline-specific writing courses can then provide opportunities to apply these structures in authentic research genres, such as theses and journal papers. Crucially, the observation that non-Linguistics students displayed considerable syntactic sophistication suggests that writing support should not be limited to language-related disciplines. A cross-disciplinary writing support framework, sensitive to both general academic writing skills and field-specific conventions, would better address students' diverse needs and promote equitable access to high-quality writing instruction.

Despite its contributions, the study has several limitations that open avenues for further inquiry. First, the analysis focused exclusively on abstracts, a concise and information-dense genre that may not capture the full spectrum of syntactic variation observable in longer sections such as introductions or discussions. Future research could expand the scope to multiple sections of academic texts to provide a more

comprehensive understanding of syntactic development across genres. Second, the classification of language background relied on educational histories in English-speaking institutions, which, while practical, may not perfectly represent native proficiency. Future studies could adopt more direct and verifiable indicators of nativeness, such as self-reported first language or national linguistic background, and examine additional levels of proficiency among EFL writers. Third, the corpus was drawn from a single institution, limiting the generalizability of the findings to broader EFL or EMI contexts in Vietnam and beyond. Expanding the sample to include multiple universities and ensuring balanced representation across disciplines would strengthen the robustness of cross-linguistic and disciplinary comparisons.

Ultimately, the study provides compelling empirical evidence that EFL writers can achieve advanced levels of syntactic complexity, particularly through phrasal elaboration, even when compared to expert native writers. While L1 writers continue to exhibit an advantage in clausal sophistication, the findings suggest that nativeness alone is not a definitive predictor of academic writing quality. Both language background and disciplinary context interact in shaping writers' syntactic strategies. By highlighting the analytical power of fine-grained syntactic measures, this study offers a more precise and equitable perspective on assessing academic writing. It calls for pedagogical approaches that explicitly integrate syntactic awareness across disciplines and for continued research into how EFL writers develop syntactic expertise in diverse academic environments. Through such efforts, writing instruction can more effectively support learners in achieving the syntactic and rhetorical sophistication demanded in global academic discourse.

ACKNOWLEDGEMENTS

This research is funded by Viet Nam National University Ho Chi Minh City (VNUHCM) under grant number C2025-28-05.

LIST OF ABBREVIATIONS

BA: Business Administration
 CS: Computer Science
 EFL: English as a Foreign Language
 EL: English Linguistics
 EMI: English-Medium Instruction
 ENL: English as a Native Language
 IE: Industrial Engineering and Systems
 L1: First Language
 L2: Second Language
 RQ: Research Question

TAASSC: The Automatic Analysis of Syntactic Sophistication and Complexity

COMPETING INTERESTS

This manuscript has no conflicts of interest.

AUTHOR'S CONTRIBUTIONS

First author: Conceived the research idea; collected and processed the data; and drafted the manuscript. Second/Corresponding author: Collected part of the data, revised and edited the manuscript, and provided funding support for the study. Scientific contribution: This study provides empirical evidence that EFL writers can attain advanced syntactic complexity comparable to native experts, highlighting that academic writing quality depends on the interaction of language background and disciplinary context rather than nativeness alone. It advances a more precise and equitable approach to assessing writing through the fine-grained syntactic analysis.

REFERENCES

- Housen A, Kuiken F. Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*. 2009;30(4):461–73. Available from: <https://doi.org/10.1093/applin/amp048>.
- Ortega L. Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*. 2003;24(4):492–518. Available from: <https://doi.org/10.1093/applin/24.4.492>.
- Casal JE, Lee JJ. Syntactic complexity and writing quality in assessed first-year L2 writing. *Journal of Second Language Writing*. 2019;44:51–62. Available from: <https://doi.org/10.1016/j.jslw.2019.03.005>.
- Jagaiah T, Olinghouse NG, Kearns DM. Syntactic complexity measures: variation by genre, grade-level, students' writing abilities, and writing quality. *Reading and Writing*. 2020;33(10):2577–638. Available from: <https://doi.org/10.1007/s11145-020-10057-x>.
- Zhang X, Lu X. Revisiting the predictive power of traditional vs. fine-grained syntactic complexity indices for L2 writing quality: the case of two genres. *Assessing Writing*. 2022;51. Available from: <https://doi.org/10.1016/j.asw.2021.100597>.
- Viera RT. Syntactic complexity in journal research article abstracts written in English. 2022;46(2). Available from: <https://doi.org/10.1016/j.jslw.2019.100666>.
- Jiang J, Bi P, Liu H. Syntactic complexity development in the writings of EFL learners: insights from a dependency syntactically-annotated corpus. *Journal of Second Language Writing*. 2019;46. Available from: <https://doi.org/10.1016/j.jslw.2019.100666>.
- Anjarwati D, Basthomi Y, Zen EL. Comparative analysis of syntactic complexity in Indonesian graduate-level writing: a corpus study of Master's theses versus Doctoral dissertations. *English Review. Journal of Engineering Education*. 2024;12(2):811–20.
- Casal JE, Lu X, Qiu X, Wang Y, Zhang G. Syntactic complexity across academic research article part-genres: A cross-disciplinary perspective. *Journal of English for Academic Purposes*. 2021;52. Available from: <https://doi.org/10.1016/j.jeap.2021.100996>.
- Dong J, Wang H, Buckingham L. Mapping out the disciplinary variation of syntactic complexity in student academic writing. *System*. 2023;113. Available from: <https://doi.org/10.1016/j.system.2022.102974>.
- Lu X, Ai H. Syntactic complexity in college-level English writing: differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*. 2015;29:16–27. Available from: <https://doi.org/10.1016/j.jslw.2015.06.003>.
- Kuiken F, Vedder I. Syntactic complexity across proficiency and languages: L2 and L1 writing in Dutch, Italian and Spanish. *International Journal of Applied Linguistics*. 2019;29(2):192–210. Available from: <https://doi.org/10.1111/ijal.12256>.
- Flowerdew J, Wang SH. Identity in academic discourse. *Annual Review of Applied Linguistics*. 2015;35:81–99. Available from: <https://doi.org/10.1017/S026719051400021X>.
- Hoang TVY, Ma LPF. Experiences of publishing in English: Vietnamese doctoral students' challenges and strategies. *Across the Disciplines*. 2019;16(3):50–66.
- Jitpraneechai N. Noun phrase complexity in academic writing: A comparison of argumentative English essays written by Thai and native English university students. *LEARN Journal: Language Education and Acquisition Research Network*. 2019;12(1):71–88.
- Esfandiari R, Ahmadi M. Phraseological complexity and academic writing proficiency in abstracts authored by student and expert writers. *English Teaching and Learning*. 2023;47(4):429–48.
- Zhou W, Li Z, Lu X. Syntactic complexity features of science research article introductions: rhetorical-functional and disciplinary variation perspectives. *Journal of English for Academic Purposes*. 2023;61. Available from: <https://doi.org/10.1016/j.jeap.2022.101212>.
- Nasseri M. Is postgraduate English academic writing more clausal or phrasal? Syntactic complexification at the crossroads of genre, proficiency, and statistical modeling. *Journal of English for Academic Purposes*. 2021;49. Available from: <https://doi.org/10.1016/j.jeap.2020.100940>.
- Ziaeeian E, Golparvar SE. Fine-grained measures of syntactic complexity in the discussion section of research articles: the effect of discipline and language background. *Journal of English for Academic Purposes*. 2022;57. Available from: <https://doi.org/10.1016/j.jeap.2022.101116>.
- Asahlanee A, Aaganathan P. Phrasal syntactic complexity measures in Linguistics research articles written by Iraqi and English L1 writers. *3L: Language, Linguistics, Literature*. 2023;29(1):154–68.
- Ellis R. *Task-based language learning and teaching*. UK: Oxford University Press; 2003.
- Bulté B, Housen A. Defining and operationalising L2 complexity; 2012. Available from: <https://doi.org/10.1075/llt.32.02bul>.
- Norris JM, Ortega L. Towards an organic approach to investigating CAF in instructed SLA: the case of complexity. *Applied Linguistics*. 2009;30(4):555–78. Available from: <https://doi.org/10.1093/applin/amp044>.
- Biber D, Gray B, Staples S. Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*. 2016;37(5):639–68. Available from: <https://doi.org/10.1093/applin/amu059>.
- Kyle K, Crossley SA. Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *Modern Language Journal*. 2018;102(2):333–49. Available from: <https://doi.org/10.1111/modl.12468>.
- Martínez AC. Analysis of syntactic complexity in secondary education EFL writers at different proficiency levels. *Assessing Writing*. 2018;35:1–11. Available from: <https://doi.org/10.1016/j.asw.2017.11.002>.
- Yin S, Gao Y, Lu X. Syntactic complexity of research article part-genres: differences between emerging and expert international publication writers. *System*. 2021;97. 102427. Available from: <https://doi.org/10.1016/j.system.2020.102427>.
- Wolfe-Quintero K, Inagaki S, Kim HY. *Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity*. Honolulu: University of Hawai'i Press; 1998.

29. Peng Y, Zheng Y, Sun J, Jiang Y, Lin J, Zhang H. Modeling relationships among large-grained, fine-grained absolute syntactic complexity and assessed L2 writing quality: an SEM approach. *Assessing Writing*. 2024;61. Available from: <https://doi.org/10.1016/j.asw.2024.100875>.
30. Biber D, Gray B, Poonpon K. Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*. 2011;45(1):5–35. Available from: <https://doi.org/10.5054/tq.2011.244483>.
31. Kyle K. Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication. Georgia State University; 2016.
32. Huang Q. A comparison of measurements of syntactic complexity in L2 writing: large-grained indices and fine-grained clausal and phrasal indices. *Cross-Cultural Communication*. 2022;18(4):78–87.
33. Staples S, Egbert J, Biber D, Gray B. Academic writing development at the university level: phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication*. 2016;33(2):149–83. Available from: <https://doi.org/10.1177/0741088316631527>.
34. Lu X. A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*. 2011;45(1):36–62. Available from: <https://doi.org/10.5054/tq.2011.240859>.
35. Lu X. Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*. 2010;15(4):474–96. Available from: <https://doi.org/10.1075/ijcl.15.4.02lu>.
36. Wang M, Lowie W. Understanding advanced level academic writing on syntactic complexity. In: *Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation*; Shanghai, China: Association for Computational Linguistics; 2021. p. 452–62.
37. Abbaspour E, Mathew P. Evaluating cohesion as a predictor of writing quality: an analysis of local, global, and text-level Indices in IELTS Writing Task 2. *International Journal of Practical and Pedagogical Issues in English Education*. 2025;3(4):97–117.
38. Staples S, Reppen R. Understanding first-year L2 writing: A lexico-grammatical analysis across L1s, genres, and language ratings. *Journal of Second Language Writing*. 2016;32:17–35. Available from: <https://doi.org/10.1016/j.jslw.2016.02.002>.
39. Lan G, Zhang Q, Lucas K, Sun Y, Gao J. A corpus-based investigation on noun phrase complexity in L1 and L2 English writing. *English for Specific Purposes*. 2022;67:4–17. Available from: <https://doi.org/10.1016/j.esp.2022.02.002>.
40. Ansarifard A, Shahriari H, Pishghadam R. Phrasal complexity in academic writing: A comparison of abstracts written by graduate students and expert writers in applied linguistics. *Journal of English for Academic Purposes*. 2018;31:58–71. Available from: <https://doi.org/10.1016/j.jeap.2017.12.008>.
41. Wu X, Mauranen A, Lei L. Syntactic complexity in English as a lingua franca academic writing. *Journal of English for Academic Purposes*. 2020;43. Available from: <https://doi.org/10.1016/j.jeap.2019.100798>.
42. Khany R, Kafshgar NB. Analysing texts through their linguistic properties: A cross-disciplinary study. *Journal of Quantitative Linguistics*. 2016;23(3):278–94. Available from: <https://doi.org/10.1080/09296174.2016.1169848>.
43. Karakaya K. A corpus-based and systemic functional analysis of syntactic complexity and nominal modification in academic writing. Iowa State University; 2017.
44. Weber-Wulff D, Anohina-Naumeca A, Bjelobaba S, Foltýnek T, Guerrero-Dib J, Popoola O, et al. Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*. 2023;19(1):1–39. Available from: <https://doi.org/10.1007/s40979-023-00146-z>.
45. Walters WH. The effectiveness of software designed to detect AI-generated writing: A comparison of 16 AI text detectors. *Open Information Science*. 2024;7(1). Available from: <https://doi.org/10.1515/opis-2022-0158>.
46. Malik MA, Amjad AI. AI vs AI: How effective are Turnitin, ZeroGPT, GPTZero, and Writer AI in detecting text generated by ChatGPT, Perplexity, and Gemini? *Journal of Applied Learning and Teaching*. 2025;8(1):91–101. Available from: <https://doi.org/10.37074/jalt.2025.8.1.9>.
47. Hyland K. *Academic discourse: English in a global context*. UK: Bloomsbury Publishing; 2009.
48. Hyland K. Writing in the university: Education, knowledge and reputation. *Language Teaching*. 2013;46(1):53–70. Available from: <https://doi.org/10.1017/S0261444811000036>.
49. Biber D, Gray B, Staples S, Egbert J. Investigating grammatical complexity in L2 English writing research: linguistic description versus predictive measurement. *Journal of English for Academic Purposes*. 2021;46:432–57.
50. Thongyoi K, Poonpon K. Phrasal complexity measures as predictors of EFL university students' English academic writing proficiency. *Reflections: The SoL Journal*. 2020;27(1):44–61. Available from: <https://doi.org/10.61508/refl.v27i1.241750>.
51. Smirnova E. Clausal complexity of expert and student writing: a corpus-based analysis of papers in social sciences. *Language Learning in Higher Education*. 2022;12(2):453–75.
52. Maamujav U, Olson CB, Chung H. Syntactic and lexical features of adolescent L2 students' academic writing. *Journal of Second Language Writing*. 2021;53. Available from: <https://doi.org/10.1016/j.jslw.2021.100822>.
53. Biber D, Gray B. Grammatical change in the noun phrase: the influence of written language use. *English Language and Linguistics*. 2011;15(2):223–50. Available from: <https://doi.org/10.1017/S1360674311000025>.
54. Wang W, Duan M, Zhang H. Corpus-based development of syntactic complexity in EFL writing. *SHS Web of Conferences*. 2023;152:04001. Available from: <https://doi.org/10.1051/shsconf/202315204001>.
55. Jones S, Myhill D, Bailey T. Grammar for writing? An investigation of the effects of contextualised grammar teaching on students' writing. *Reading and Writing*. 2013;26(8):1241–63. Available from: <https://doi.org/10.1007/s11145-012-9416-1>.

Độ phức tạp cú pháp cấp độ tinh trong các bản tóm tắt nghiên cứu: nghiên cứu so sánh giữa sinh viên Việt Nam và các tác giả bản ngữ chuyên gia theo ngành học

Bùi Biện Thảo Nhi, Nguyễn Thị Quyên*



Use your smartphone to scan this QR code and download this article

TÓM TẮT

Độ phức tạp cú pháp, thể hiện qua các câu dài và nhiều mệnh đề lồng nhau, là một đặc điểm quan trọng của văn bản học thuật chất lượng vì nó cho phép người viết diễn đạt các ý tưởng phức tạp một cách súc tích và hiệu quả. Mặc dù nhiều nghiên cứu trước đây đã xem xét độ phức tạp cú pháp ở mức độ khái quát, các nghiên cứu sử dụng chỉ số đo chi tiết vẫn còn hạn chế. Nghiên cứu này khảo sát độ phức tạp cú pháp cấp độ tinh trong các bản tóm tắt nghiên cứu do sinh viên Việt Nam viết và so sánh với các bản tóm tắt của các tác giả bản ngữ tiếng Anh có trình độ chuyên gia. Bên cạnh đó, nghiên cứu cũng xem xét sự khác biệt giữa sinh viên ngành Ngôn ngữ Anh và sinh viên các ngành ngoài Ngôn ngữ Anh. Kết quả phân tích cho thấy có sự khác biệt rõ ràng giữa sinh viên và chuyên gia ở cả các chỉ số mệnh đề và cụm từ, đồng thời xuất hiện sự khác biệt đáng kể giữa các lĩnh vực chuyên môn. Nhìn chung, các bản tóm tắt của sinh viên thể hiện mức độ phức tạp cú pháp tiệm cận với các tác giả chuyên gia, mặc dù sinh viên ngành Ngôn ngữ Anh sử dụng cấu trúc cú pháp kém phức tạp hơn so với các bạn thuộc nhóm ngành khác. Những phát hiện này góp phần làm sáng tỏ ảnh hưởng của yếu tố ngành học và nền tảng ngôn ngữ đối với việc viết học thuật, đồng thời gợi mở những hàm ý cho lý thuyết và giảng dạy.

Từ khoá: độ phức tạp cú pháp cấp độ tinh, khác biệt theo ngành học, nền tảng ngôn ngữ, tác giả bản ngữ chuyên gia, bản tóm tắt nghiên cứu

Trường Đại học Quốc tế, ĐHQG-HCM, Việt Nam

Liên hệ

Nguyễn Thị Quyên, Trường Đại học Quốc tế, ĐHQG-HCM, Việt Nam

Email: ntquyen@hcmu.edu.vn

Lịch sử

- Ngày nhận: 14-10-2025
- Ngày sửa đổi: 06-02-2026
- Ngày chấp nhận: 25-05-2026
- Ngày đăng: 28-09-2026

DOI : <https://doi.org/10.32508/vnuhcmjssh.v10i3.1257>



Bản quyền

© Tạp chí ĐHQG-HCM. Đây là bài báo công bố mở được phát hành theo các điều khoản của the Creative Commons Attribution4.0 International license.

Trích dẫn bài báo này: Nhi B B T, Quyên N T. **Độ phức tạp cú pháp cấp độ tinh trong các bản tóm tắt nghiên cứu: nghiên cứu so sánh giữa sinh viên Việt Nam và các tác giả bản ngữ chuyên gia theo ngành học.** *VNUHCMJ. Soc. Sci. Humanit.* 2026;10(3):4127-4143.