A Corpus-Based Study on Phrasal Complexity in Computer Science Abstracts of Novice and Advanced Writers

Qing Xue, Tianshuang Ge[*]

Department of Foreign Languages, Dalian University of Technology, Dalian, China
Email: *tianshuangge@dlut.edu.cn

Abstract

The correlation between phrasal complexity and L2 writers' writing proficiency has been confirmed by numerous studies, among which the empirical studies on measuring of L2 writing have attracted much attention in English for academic purposes. Based on the hypothesized developmental indices of noun phrase modifications, this study examines, from the corpus-driven approach, the linguistic realization of phrasal complexity on abstracts written by international advanced academic writers and Chinese graduate L2 writers of computer science. The results show that while both groups illustrated the developmental stages of academic writing, many complexity measures of noun phrase modifiers significantly distinguish L2 learners from expert writers, especially in post-modifiers. The findings thus support previous studies on writing developmental stages, presenting a thorough understanding of graduate L2 writers' and advanced writers' of EAP writing through analysis of pre-modifiers and post-modifiers.

Keywords

Phrasal Complexity, Complex Noun Phrases, L2 Academic Writing, Computer Science

1. Introduction

Conventional wisdom on abstracts demonstrates that an abstract contains a crucially important content of an already existing text (Kilborn, 1998). As an indispensable screening and indexing tool (Huckin, 2001), free access to complete full-text of research article abstracts blurs or breaks the boundaries of interactive exchange between academic communities (Koltay, 2010). Concerning abstract composing, it not only lies in the writing strategies and linguistic devices that

urge writers to compose a remarkably compact text with relatively few words, but also brief, concise, and objective previews that assist readers in gaining convenient access to information. For instance, Ruan (2018) has pointed out that constructing such highly condensed texts requires authors to wield a sophisticated repertoire of linguistic and non-linguistic knowledge.

Additionally, evidence has shown that there is a notable historical surge in the use of "nouns and phrasal noun-modifying structures" in academic writing (Biber & Gray, 2016). In particular, research article abstracts as a unique genre in academic prose, contain a large number of lengthy noun phrases and non-clausal phrases (e.g., Biber & Gray, 2010; Biber et al., 2011; Ansarifar et al., 2018; Ruan, 2018; Song & Wang, 2019), which present a highly compact writing style. Given these circumstances, it is natural to wonder what features are presented in use; an exploration of noun phrases in research article abstracts will thus contribute to understanding the developmental stages of complexity features in academic writing.

Although it is commonly thought that an attractive abstract is crucial to a wide range of academic writers, several issues involved are still open to question. First, studies have shown that there is indeed a poor competence of structural clarity and expressive accuracy for second language (L2) academic writers, especially abstracts of hard sciences (Fu et al., 2021), which affects journal visibility, citation frequency (Wang, 2020), as well as author's academic identity construction. Second, owing to lack of detailed guidance and targeted feedback, these novices writers, more specifically in non-native bilingual or multilingual writers, are often capable of constructing their abstracts with "maximum efficiency, clarity, and economy" (Swales & Feak, 2009). Consequently, as for abstract constructing instruction, it should be an essential and integral component of teaching objectives to effectively identify novice L2 writers' developmental stages and provide targeted feedback and guidance according to the differences in phrasal complexity.

Moreover, despite many scholars indicate that abstract is sensitive enough to distinguish research article abstracts across various disciplines (e.g., Hyland & Tse, 2004; Pho, 2008; Omidian et al., 2018), what remains largely unknown is what rhetorical strategies are selected, what fundamental distinctions are presented, and how developmental stages are obeyed, especially within a concrete discipline. That is to say, an intermediate investigation into academic sub-registers still needs to attach due significance to recognize the distinctive patterns of employed linguistic recourses (Biber & Gray, 2016: p. 250). However, there are few studies so far probing into the grammatical devices concerning the realization of structural compression in scientific research writing. If the answers to these questions are pretty ambiguous and fail to response undoubtedly, then targeted instruction in constructing abstracts might become no more than an exhilarating proposal.

This study aims to contribute instructions on language-focused EAP (English for Academic Purposes) classrooms and teaching reform in advanced academic

writing to serve effective academic communication by comparing novice Chinese writers and more advanced international writers. Rather than simply present speculative interpretations of the overall writing development, this study uses hypothesis of Biber et al.'s (2011) to empirically investigate specific stages of writing development on computer science that is traditionally regarded as hard sciences. The concrete group of L2 writers studying here is Chinese graduates who have published their dissertations in CNKI (China National Knowledge Infrastructure) but lack extensive publication experience in international academic journals. We compare the phrasal complexity of research articles published in CNKI by novice writers and those published in IEEE INFOCOM (IEEE International Conference on Computer Communication), a leading conference in the same field, by advanced writers with many publications in preeminent journals. It is assumed that abstracts written by top-tier researchers are regarded as the model of composing influential condensed texts, and features of advanced writers' writing would then have practical implications for novice and less experienced L2 academic writers.

In light of these dynamics, this study mainly aims to answer the following questions:

1) Whether novice Chinese writers and advanced international writers follow the development stages for noun phrase modification features hypothesized by Biber et al. (2011)?

2) What are the similarities and distinctions of the linguistic features of noun phrases between novice L2 writers and advanced international writers in computer science?

3) What are the reasons behind the differences, and what implications for targeted language-focused instruction for Chinese graduate computer science writers?

This paper first presents an overview of syntactic complexity in writing quality and development, followed by the more recent studies focusing on phrasal complexity indices. It then summarizes the developmental stages for phrasal complexity features from developmental stages for complexity features (Biber & Gray, 2010). Finally, the paper focuses on phrasal complexity indices to predict the writing quality and developmental stages of Chinese L2 graduates, hoping to have a thorough understanding of the fundamental distinctions between novice writers and advanced writers.

2. Measuring Linguistic Complexity in Second Language Writing

As a sub-dimension of L2 English writing (Wolfe-Quintero et al., 1998; Ortega, 2003), linguistic complexity studies (also called syntactic complexity studies) has long sought ways to explore what language features found within text could effectively assess both writing development and language proficiency of L1 and L2 writers (Pallotti, 2015; Lu, 2017; Kim & Crossley, 2018; Casal & Lee, 2019; Bi &

Jiang, 2020). An early study on complexity measurement is performed through a great number of "global measures" (Ortega, 2003), including the amount of embedding, the sophistication of structure, the range of structural types, as well as the length of production unit.

More recently, however, the conventional or "large-grained" length and clausal-based metrics (Kyle & Crossley, 2018) have been greeted with some skepticism (Biber et al., 2011; Kyle & Crossley, 2018; Biber et al., 2020; Crossley, 2020). One of the two major concerns is that excessive attention at the syntactic level is deemed unnecessary to a greater or lesser extent (Kyle & Crossley, 2018; Biber et al., 2020). On the one hand, findings from the historical change in spoken and written discourse demonstrate that complex noun phrases rather than clausal subordination can explicitly delineate the boundaries between written academic prose (especially academic research writing) (Biber & Gray, 2010; Biber et al., 2011). On the other hand, a large number of studies primarily take syntactical units (e.g., mean length of T-unit) as complexity indices; however, no in-depth attempt has been made to operationalize linguistic complexity at the phrase level. The underlying notion is that the length of a T-unit may increase systematically, both from the extensive use of phrasal dependents (such as attributive adjectives) to the extensive use of dependent clauses (such as finite complement clauses); however, large-grained indices fail to capture the structural devices deployed within a T-unit (Kyle & Crossley, 2018). The other inexplicable question on large-grained measures is that they do not afford accurate, adequate, and satisfactory explanations for "additional layer of meaning" (Paquot, 2019) manifested in specific lexical patterns (Kyle & Crossley, 2018; Biber et al., 2020), as they show weakness in drawing distinctions between phrasal and clausal complexity (Biber et al., 2020). Consequently, Biber et al., 2011, 2020, Biber & Gray, 2016 have proposed that L2 writing assessment should pay more attention to the complexity indices at the phrasal level in academic writing, thus constructing a hypothesis of five writing developmental stages from preference for clauses to a dense use of noun phrases.

Therefore, the author's contention here is that the large-grained indices on syntactical complexity are somewhat controversial concerning its interpretive power in accounting for phrasal complexity features, thus requiring more fine-grained measures of phrasal complexity in a specific disciplinary field (computer science) than it has hitherto received.

3. Methodology

3.1. Theoretical Framework

It is widely recognized that syntactic complexity is associated with greater wiring development. Traditional indices of complexity shed lights on clause-level structure, ranging from T-unit counts to sentence length (e.g., Ortega, 2003; Lu, 2011). Supported by a comparative study on spoken and written discourse, Biber, Gray and Poonpon (2011) have challenged this stereotype on complexity in

English academic prose by redefining it as more elaboration (clausal complexity) acquired in early developmental stages and more compression (phrasal complexity) in later stages. Specifically, Biber et al. (2011) divides writing development into five stages. The stages progress from "finite dependent clauses functioning as constituents in other clauses", toward intermediate stages of "nonfinite dependent clauses and phrases functioning as constituents in other clauses", and finally to the last stage requiring "condensed use of phrasal (non-clausal) dependent structures that function as constituents in noun phrases" (Biber et al., 2011).

More recent studies are further pushing writing development measurement from clausal complexity and moving toward phrasal complexity under the presumption that high-quality academic writing is closely associated with a higher degree of phrasal complexity (Biber et al., 2011; Lu, 2011; Taguchi et al., 2013; Parkinson & Musgrave, 2014), while at the same time posing a comparative absence of clause or sentence devices. Biber (1988: p. 104) has demonstrated that the high frequency of nouns as the fundamental carriers of referential meaning is linked to a higher information density. As the drift towards a structurally compressed discourse style in academic written registers (Biber & Gray, 2016), it turns out that abstracts present an extremely dense use of phrasal structures in limited word count where information is conveyed through phrasal devices instead of clausal elaboration (Biber & Gray, 2010; Ansarifar et al., 2018; Yin et al., 2021).

Furthermore, multiple studies have confirmed Biber's claims about the significance of phrase-level structure as more convincing measures of complexities of English academic writing (e.g., Lu, 2011; Kyle & Crossley, 2018; Ruan, 2018). Corresponding studies have explored L2 writing development by comparing writing in different genres such as argumentative writing, course essay, and critiques (e.g., Lu, 2011; Taguchi et al., 2013; Atak & Saricaoglu, 2021) or by comparing L2 writers at different levels such as secondary, undergraduate, and graduate level (e.g., Casal & Lee, 2019; Bi & Jiang, 2020; Yin et al., 2021). The following results from such research confirm the statement that phrase-level indices perform better than clause-level indices in L2 writing development. It thus provides a strong starting point for discussing considerations into phrase-level complexity indices in general.

Therefore, the current study builds on the developmental stages for complexity features hypothesized by Biber et al. (2011), which jointly utilize the pre- and post-modifiers embedded to measure L2 writing quality and development. Table 1 presents the developmental stages for noun phrase features, which have been continuously improved and enriched according to the research object of this study to make it more in line with the present research purpose. For instance, in the original framework, simple prepositional phrases with concrete and abstract meaning were in the third and fourth developmental stages respectively. However, considering the discipline characteristics of computer science, there is no significant difference between advanced and novice writers, as both of them

frequently use simple prepositional phrases with abstract meaning (e.g., electronic communication signals, 5G networks, Internet of Things). Additionally, there are no clear criteria to classify abstract or concrete sense in these novel fields. Therefore, simple prepositional phrases with concrete and abstract meaning were separated in the study.

As shown in Table 1, the indices related to noun phrases are not presented in stage 1, and start directly from stage 2 through simple phrasal embedding in the noun phrase and prepositional phrases with relative clauses; further developments include modifying nouns with -ed and -ing clauses and modifying nouns with prepositional phrases; the final stage reflected in Table 1 includes phrasal structures as noun modifiers, and complementary clauses as noun modifiers.

3.2. Corpus

To carry out a contrastive study of abstracts written by advanced academic writers and novice graduate writers of computer science, 400 relevant abstracts are collected from two sources covering the period between 2019 and 2020 (Table 2). One is Chinese graduate writers' abstracts from CNKI (https://www.cnki.net/), and the other is advanced academic writers' abstracts from IEEE INFOCOM (https://infocom2020.ieee-infocom.org/). During data collection, each text is stored in plain TXT files and checked manually to ensure that the primary topic is closely relevant to computer science. Additionally, the original data of the research is stored in a separate Excel file that contains the information on authors, journals, word counts, web addresses, publishing and collection dates so that the original files and information can be retrieved.

Table 1. Developmental stages for noun phrase features.

Stage	Noun phrase structures	Example from this study
2	Attributive adjectives	wearable sensors; unique physiological characteristics
3	Participial premodifiers	matching accuracy; implemented system
	That relative clauses, especially with animate head nouns	a recognition system that can work under different usage scenarios
	Nouns as premodifiers	system robustness
	Possessive nouns	network's capabilities
	Of phrases	the robustness of MagView
	Prepositions other than of	broad applications in the Internet
4	Nonfinite relative clauses(-ed)	groups of items carried by a customer
	Nonfinite relative clauses(-ing)	an SLAM service running on mobile devices
	Multiple modifiers of Attributive adjectives and nouns	an efficient computation offloading strategy
5	Preposition + nonfinite complement clause	an AM station for broadcasting short messages
	Complement clauses controlled by nouns	the fact that the covert channel can be created on any computer
	Multiple prepositional phrases, with levels of embedding	the confusion between the two concepts of scientific knowledge map in the field of library and information and Google Knowledge Graph
	To-clauses	the right to reliable meteorological data

Table 2. Description of the corpus of abstracts.

Sub-corpus	Data sources	No of texts	No of words	Mean length
IEEE writers' abstracts (IWA)	IEEE International Conference on Computer Communication	200	36,848	184.24
Chinese writers' abstracts (CWA)	Master's Theses from CNKI	200	79,373	396.87

CNKI has built the most comprehensive system of academic resources in Mainland China, serving as a window to research on Chinese graduates with a Master's Degree in computer science. Dissertations are published in Chinese, accompanying with English research abstracts. IEEE INFOCOM, describing significant and innovative research contributions to communications networks, is selected as a cutting-edge conference on networking in the research community recognized by academia worldwide. It is a major conference venue for researchers and scholars with a variety of linguacultural backgrounds to present and exchange significant and innovative contributions and ideas in the field of networking and closely related areas, covering both theoretical and systems research. Additionally, the decision on selecting the IEEE INFOCOM was based on the assumption that abstracts written by top-tier researchers and advanced writers are regarded as the model of composing influential condensed texts, and features of specialist academic writing would then have practical implications for novice and less experienced L2 academic writers.

3.3. Analytical Procedures

The two corpora were tagged using CLAWS part-of-speech tagger (Garside & Smith, 1997), and manual coding fixing was conducted for noun phrase structures that couldn't be tagged automatically. For instance, each instance with Of phrases was checked manually to further differentiate whether it was a preposition with nonfinite complement clause in stage 5 (e.g., a prototype of using low cost …) or a noun post-modification in stage 3 (e.g., localization of illegal transmitters). Additionally, AntConc (Anthony, 2019) was applied to identify the frequency of noun phrase structures in the two corpora by retrieving the corresponding tag of the indices presented in Table 1, aiming to explore the similarities and differences in the use of noun phrases. Moreover, the independent-samples T-test is employed as a method to explore significant differences between advanced international writers' and novice Chinese writers' abstracts. Finally, this study shed lights on the reasons behind the current stages and hopes to offer some suggestions for EAP writing.

4. Results

4.1. Predicting and Assessing L2 Writing Development

In order to investigate whether abstracts written by novice L2 writers and advanced writers follow the developmental writing stages hypothesized by Biber et al. (2011), in-depth analyses on five developmental stages were applied. A large

number of studies have confirmed the importance of complex nouns as a complexity measure to distinguish learners from expert writers (e.g., Ruan, 2018; Larsson & Kaatari, 2020; Bychkovska, 2021); thus, 14 indices closely related to complex nominals presented in Table 1 are utilized to predict writing quality and development of novice and advanced writers.

Figure 1 below displays the proportion of noun phrase structure deployed in four developmental stages. As can be seen, noun phrase features from stage 3 are the most frequent linguistic resources for constructing abstracts for both novice and advanced writers, at 45.64% and 45.07% respectively. Additionally, an interesting trend is that the proportion of two corpora in other stages also exposes remarkable similarity (e.g., 27.85% and 29.92 at stage 4, 21.63% and 23.68% at stage 2), demonstrating that Chinese novice writers possess analogous preferences with advanced writers in writing strategies and language resources of composing abstracts. These similarities may result from writers' perspective of academic identity construction and academic writing exploration wherein for novice L2 writers they favor writing strategies that manifested conform to expected norms and authorities.

However, although noun phrase structures from both corpora presented least at stage 5, a more significant distinction is displayed between novice and advanced writers at 4.88% and 1.33% respectively. Moreover, compared with advanced international writers, novice Chinese writers have a relatively higher writing complexity at stage 2, with a proportion of 23.68%. These findings are broadly in line with the previous studies, that is, higher quality and more advanced writing involves more syntactic complexity (Kyle & Crossley, 2018; Ansarifar et al., 2018; Crossley, 2020; Atak & Saricaoglu, 2021). Furthermore, the proportion of various stages in Table 3 may indicate that IEEE writers, as an advanced level of English learners or English natives, utilize greater linguistic recourses characterizing more sophisticated developmental stages, while Chinese graduates, as Chinese English learners at a relatively high level, still show quite reliance on features at the early stages of L2 development, and pose weakness in an advanced combination of various linguistic skills.

To conclude, without sufficient empirical proof for the claim that the top-ranked research articles are salient enough to enlighten academic writing of novice L2 writers, it is hard to envisage such expected assumptions; therefore, a further investigation is carried out into complex nominals of the novice and advanced academic writing.

4.2. Novice vs. Advanced Academic Writing in Computer Science

Normalized frequency analysis is conducted in an attempt to explore various types of modifiers included in complex nominals. It does so by examining whether there are apparent or observable differences in the frequencies of pre- and post-modifications between the two corpora, followed by independent-samples T-test of their statistical significances through IBM SPSS Statistics (v.25).
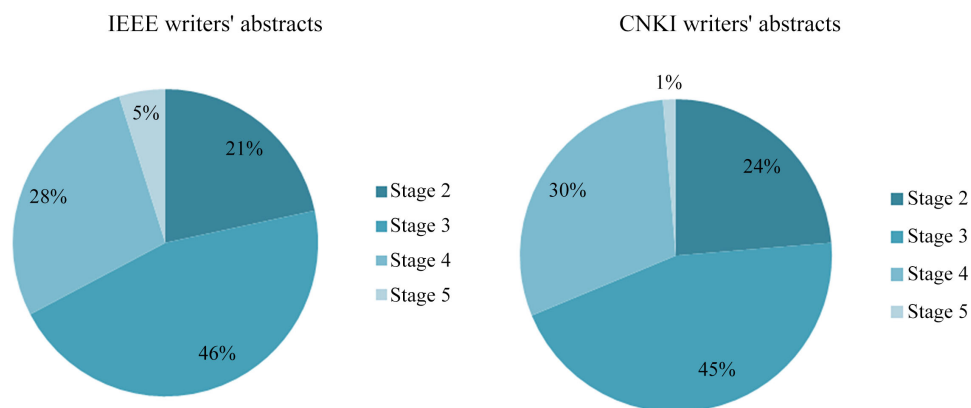
IEEE writers' abstracts   CNKI writers' abstracts

Figure 1. Percentage of noun phrase structures in four developmental stages.

Table 3. Noun phrase structures in four developmental stages.

Stage	IEEE writers' abstracts (IWA)			CNKI writers' abstracts (CWA)		
	Raw freq.	Per 100 words	Percent	Raw freq.	Per 100 words	Percent
2	1872	5.08	21.63%	3869	4.87	23.68%
3	3949	10.72	45.64%	7366	9.28	45.07%
4	2410	6.54	27.85%	4890	6.16	29.92%
5	422	1.15	4.88%	217	0.27	1.33%
Total	8653	23.48	100.00%	16342	20.59	100.00%

Figure 2 graphically illustrates the distribution of phrasal features. Considering the total words in IWA are more than twice those in CWA, raw frequency is standardized per 100 words to ensure the controllable comparison between the two corpora. As shown in Figure 2, the most common types of noun post-modifying noun are structure with prepositional phrases, while attributive adjectives and nouns were the most common form of pre-modifying nouns. It provides supporting evidence for widespread usage of attributive adjectives and prepositional phrases in written academic registers (Biber et al., 2020) while showing an inconsistent claim that science research writing contains fewer attributive adjectives (Biber & Gray, 2016). Furthermore, the finding demonstrates that computer science as a hard science discipline also shows somewhat similarity to applied linguistics (Ansarifar et al., 2018), traditionally regarded as a humanistic discipline in phrasal structures.

Table 4 represents raw frequencies, normalized frequencies, and percentages of each measure in the two corpora. It demonstrates that all the factors concerning post-modifying nouns are differed between the two groups of writers, from usages of prepositions at an earlier stage to multiple prepositional phrases with levels of embedding at a more advanced stage. For instance, advanced writers show a preference for prepositional phrases other than of, while Chinese writers tend to use more NP-of phrases. Ansarifar et al.'s (2018) partially confirmed this finding that MA writers are prone to underuse prepositional noun modifiers compared to expert writers. Additionally, a further investigation into the NP-of phrases manifests that Chinese writers treat NP-of phrases as part of

possessive case governed by the study object of the thesis (e.g., of paper, of work, of system, of an algorithm). On the other hand, the advanced writers more frequently use prepositional phrases in a cause-effect relationship, thus increasing the complexity of the internal logical structure of abstracts through the compression of additional information by post-modification.
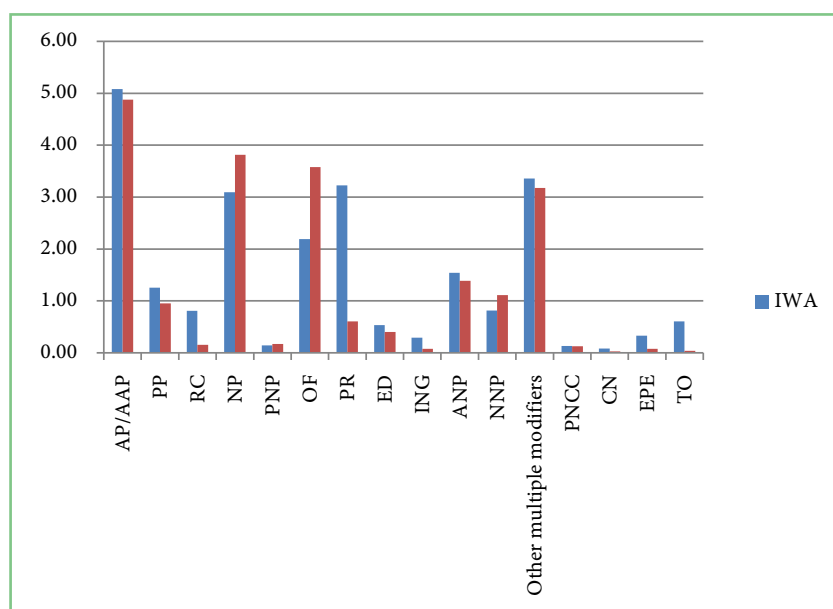

Figure 2. The normalized frequency rates per 100 words.

Table 4. Frequencies of each modifier in two sub-corpora.

Stages	Grammatical features	Modifiers	IEEE writers' abstracts (IWA)			CNKI writers' abstracts (CWA)		
			Raw freq.	Per 100 words	Percent	Raw freq.	Per 100 words	Percent
2	Attributive adjectives	AP/AAP	1872	5.08	21.63%	3869	4.87	23.68%
3	Participial premodifiers	PP	463	1.26	5.35%	756	0.95	4.63%
	That relative clauses, especially with animate head nouns	RC	299	0.81	3.46%	124	0.16	0.76%
	Nouns as premodifiers	NP	1140	3.09	13.17%	3029	3.82	18.54%
	Possessive nouns	PNP	52	0.14	0.60%	135	0.17	0.83%
	Of phrases	OF	807	2.19	9.33%	2842	3.58	17.39%
	Simple prepositional phrases other than of	PR	1188	3.22	13.73%	480	0.60	2.94%
4	Nonfinite relative clauses: -ed	ED	196	0.53	2.27%	320	0.40	1.96%
	Nonfinite relative clauses: -ing	ING	107	0.29	1.24%	60	0.08	0.37%
	Multiple modifiers of Attributive adjectives and nouns	ANP	568	1.54	6.56%	1103	1.39	6.75%
		NNP	301	0.82	3.48%	884	1.11	5.41%
		Other multiple modifiers	1238	3.36	14.31%	2523	3.18	15.44%
5	Preposition + Nonfinite complement clause	PNCC	48	0.13	0.55%	102	0.13	0.62%
	Complement clauses controlled by nouns	CN	30	0.08	0.35%	22	0.03	0.13%
	Multiple prepositional phrases, with levels of embedding	EPE	121	0.33	1.40%	61	0.08	0.37%
	To-clauses	TO	223	0.61	2.58%	32	0.04	0.20%

Table 5 below displays the comparisons of the normalized occurrences of each type of modifiers in the two sub-corpora. Mean occurrences of all phrasal complexity measures were compared by employing independent-samples T-test for statistical significance ($p < 0.05$). Overall, it is observed that all significant differences between the two corpora appear in post-modifying nouns. Of the seventeen measures, more considerable variation was observed in stages 3 and 5 than in other stages. In stage 2, no significant between attributive adjectives was found for either Chinese graduate writers or advanced international writers; one possible reason may be that attributive adjectives are linguistic devices acquired at an earlier stage of L2 acquisition. In stage 3, that relative clauses and prepositional phrases other than of were found to have statistically significant differences, accounting for a p-value of 0.001 and 0.000, respectively. Furthermore, although scholars have hypothesized that participle adjectives are acquired later than other pre-modifiers (Parkinson & Musgrave, 2014), no significant differences between advanced and novice writer corpora were found in this study. A review of stage 5 also shows that more advanced writers tend to use more post-modifying nouns through the higher frequent embedding of to-clauses in their academic writing, with statistically significant differences found between the IWA and the CWA groups.

Table 5. Comparisons of occurrences of modifiers after normalization (per 100 words).

Stage	Grammatical feature	Modifiers	Abstracts of IEEE writers (N = 50)		Abstracts of CNKI writers (N = 50)		Mean difference
			Mean	SD	Mean	SD	P(t-test)
2	Attributive adjectives	AP	4.43	2.01	4.19	2.27	0.577
		AAP	0.65	0.57	0.69	0.9	0.791
3	Participial premodifiers	PP	1.26	0.97	0.95	0.89	0.099
	That relative clauses, especially with animate head nouns	RC	0.81	1.1	0.16	0.86	0.001*
	Nouns as premodifiers	NP	3.09	2.27	3.82	1.79	0.077
	Possessive nouns	PNP	0.14	0.4	0.17	0.43	0.719
	Of phrases	OF	2.19	1.99	3.58	6.33	0.142
	Simple prepositional phrases other than of	PR	3.22	2.79	0.6	2.35	0.000*
4	Nonfinite relative clauses: -ed	ED	0.53	1.04	0.4	1.29	0.58
	Nonfinite relative clauses: -ing	ING	0.29	1.01	0.08	0.73	0.236
		ANP	1.54	1.29	1.39	1.17	0.544
	Multiple modifiers of Attributive adjectives and nouns	NNP	0.82	0.81	1.11	1.39	0.205
		Other multiple modifiers	3.36	1.65	3.18	1.78	0.601
5	Preposition + Nonfinite complement clause	PNCC	0.13	0.43	0.13	0.87	0.942
	Complement clauses controlled by nouns	CN	0.08	0.2	0.03	0.34	0.372
	Multiple prepositional phrases, with levels of embedding	EPE	0.33	1.24	0.08	1.25	0.318
	To-clauses	TO	0.61	0.94	0.04	0.43	0.000*

* $p < 0.002$.

5. Discussion

This study conduct a corpus-driven comparative studies on abstracts written by international advanced writers and Chinese novice L2 writers of computer science and found that both group of scientific research writers follow the development stages for noun phrase modification features hypothesized by Biber et al. (2011). There results support findings from previous studies that the increasing complexity development and the raising-quality academic writing present greater reliance on noun phrases or phrasal complexity (e.g., Biber et al., 2011; Kyle & Crossley, 2018; Ansarifar et al., 2018; Biber et al., 2020; Crossley, 2020; Larsson & Kaatari, 2020; Atak & Saricaoglu, 2021; Bychkovska, 2021).

To conclude, there are several implications of this study for L2 phrasal complexity. First, as for Chinese graduate writers, the disproportionate focus on pre-modifying nouns and relative insufficiency of post-modifying nouns might deserve more attention. On the one hand, these issues are closely related to academic writing norms, writer identity construction, and international academic discourse status, all of which influence the students' academic writing practice. On the other hand, these academic writing skills are expected to be acquired by students at the graduate level when dealing with academic studies. Still, many of them fail to deploy linguistic strategies to construct compact writing tactically. The underlying notion is that academic writing for Chinese writers is supposed to move beyond blindly or mechanically embedding multiple modifiers, but attempt to raise awareness on shift to an integrated combination of noun phrase features concerning more advanced stages. Specifically, as advanced writers demonstrate movement toward more complex phrasal structure, Chinese graduate writers are supposed to push the inherent notion of academic writing away from the overuse of simple noun phrases at early stages (e.g., attributive adjectives, nouns, and NP-of phrases), and moving more toward producing complex non-clausal phrases and lengthy post-modifying nouns at later stages (e.g., prepositional phrases other than of, and multiple prepositional phrases with levels of embedding). Furthermore, an exploration into abstracts in computer science also sheds new light on disciplinary variations of the compact discourse style in academic prose, affording explicit guidance for L2 writers on what phrasal features are able to use and what features are supposed to focus in academic writing of computer science. For instance, given the preciseness and accuracy of research articles in computer science, it should be noted that science research writing contains fewer attributive adjectives (Biber & Gray, 2016). Besides, although computer science (hard science discipline) demonstrates a somewhat similar phrasal structure to applied linguistics (humanistic discipline) (Ansarifar et al., 2018), academic writing of computer science is more characterized by a structurally compressed discourse style (high proportion of complex noun phrases) in academic written registers.

Second, regarding language-focused EAP classrooms and teaching reform in advanced academic writing, the findings have shown that higher-level writers

frequently use modifiers at higher developmental stages to produce condensed discourse style, and linguistic elements of phrasal complexity can be used to predict and access writing development of L2 writers. It seems reasonable to assume that corresponding instructional guidance could positively impact L2 writers' writing development. Therefore, teachers may provide extensive practice in decoding complex noun phrases in actual academic writing teaching. Furthermore, it is worth setting the teaching objective of academic prose more precisely than the large-grained instruction that misplaces the focus on a syntactic or clausal level, with more explicit guidance and feedback on phrasal complexity under which complex nouns should be realized, grasped, and utilized by students. More studies should be carried on to discuss the impact of shorter-term or longer-term interventions on student's writing later in the following study. Additionally, this study provides a basis for constructing complex noun phrases or compressed discourse styles in other disciplines. It also offers a strong starting point for setting up following-up studies into further corpus expansion, instructional textbooks, and formative teaching in general.

6. Conclusion

In conclusion, these lines of studies may advance our understanding of the writing quality and development of incorporating explicit noun phrases instruction into L2 research article writing. However, it poses a somewhat limitation that the finding of use patterns and the subsequent instructional guidance are not practically validated in real classrooms; thus it would be a productive area for future study to conduct classroom-based studies to assess the specific instructional method for phrasal complexity.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Ansarifar, A., Shahriari, H., & Pishghadam, R. (2018). 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, 31,* 58-71. https://doi.org/10.1016/j.jeap.2017.12.008

Anthony, L. (2019). *AntConc* [Computer Software]. Waseda University. http://www.laurenceanthony.net

Atak, N., & Saricaoglu, A. (2021). Syntactic Complexity in L2 Learners' Argumentative Writing: Developmental Stages and the Within-Genre Topic Effect. *Journal of English for Academic Purposes, 47,* Article ID: 100506.

https://doi.org/10.1016/j.asw.2020.100506

Bi, P., & Jiang, J. (2020). Syntactic Complexity in Assessing Young Adolescent EFL Learners' Writings: Syntactic Elaboration and Diversity. *System, 91,* Article ID: 102248. https://doi.org/10.1016/j.system.2020.102248

Biber, D. (1988). Variation across Speech and Writing. Cambridge University Press. https://doi.org/10.1017/CBO9780511621024

Biber, D., & Gray, B. (2010). Challenging Stereotypes about Academic Writing: Complexity, Elaboration, Explicitness. *Journal of English for Academic Purposes, 9,* 2-20. https://doi.org/10.1016/j.jeap.2010.01.001

Biber, D., & Gray, B. (2016). *Grammatical Complexity in Academic English, Linguistic Change in Writiing.* Cambridge University Press. https://doi.org/10.1017/CBO9780511920776

Biber, D., Gray, B., & Poonpon, K. (2011). Should We Use Characteristics of Conversation to Measure Grammatical Complexity in L2 Writing Development? *TESOL Quarterly, 45,* 5-35. https://doi.org/10.5054/tq.2011.244483

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,* Article ID: 100869. https://doi.org/10.1016/j.jeap.2020.100869

Bychkovska, T. (2021). Effects of Explicit Instruction on Noun Phrase Production in L2 Undergraduate Writing. *Journal of English for Academic Purposes, 54,* Article ID: 101040. https://doi.org/10.1016/j.jeap.2021.101040

Casal, E., & Lee, J. J. (2019). Syntactic Complexity and Writing Quality in Assessed First-Year L2 Writing. *Journal of Second Language Writing, 44,* 51-62. https://doi.org/10.1016/j.jslw.2019.03.005

Crossley, S. A. (2020). Linguistic Features in Writing Quality and Development: An Overview. *Journal of Writing Research, 11,* 415-443. https://doi.org/10.17239/jowr-2020.11.03.01

Fu, Y., Wang, H. L., & Wang, Z. Q. (2021). Scientific Paper Summarization Using Macro Discourse Structure. *Journal of Computer Applications*.

Garside, R., & Smith, N. (1997). *CLAWS Part-of-Speech Tagger for English* [Computer Software]. http://ucrel.lancs.ac.uk/claws/

Huckin, T. N. (2001). Abstracting from Abstracts. In M. Hewings (Ed.), *Academic Writing in Context* (pp. 93-103). University of Birmingham Press.

Hyland, K., & Tse, P. (2004). Metadiscourse in Academic Writing: A Reappraisal. *Applied Linguistics, 25,* 156-177. https://doi.org/10.1093/applin/25.2.156

Kilborn, J. (1998). *Writing Abstracts.* http://leo.stcloudstate.edu/bizwrite/abstracts.html

Kim, M., & Crossly, S. A. (2018). Modeling Second Language Writing Quality: A Structural Equation Investigation of Lexical, Syntactic, and Cohesive Features in Source-Based and Independent Writing. *Assessing Writing, 37,* 39-56. https://doi.org/10.1016/j.asw.2018.03.002

Koltay, T. (2010). *Abstracts and Abstracting: A Genre and Set of Skills for the Twenty-First Century.* Chandos Publishing.

Kyle, K., & Crossley, S. A. (2018). Measuring Syntactic Complexity in L2 Writing Using Fine-Grained Clausal and Phrasal Indices. *Modern Language Journal, 102,* 333-349. https://doi.org/10.1111/modl.12468

Larsson, T., & Kaatari, H. (2020). Syntactic Complexity across Registers: Investigating

(In)Formality in Second-language Writing. *Journal of English for Academic Purposes, 45,* Article ID: 100850. https://doi.org/10.1016/j.jeap.2020.100850

Lu, X. (2011). A Corpus-Based Evaluation of Syntactic Complexity Measures as Indices of College-Level ESL Writers' Language Development. *TESOL Quarterly, 45,* 36-62.

Lu, X. (2017). Automated Measurement of Syntactic Complexity in Corpus-Based L2 Writing Research and Implications for Writing Assessment. *Language Testing, 34,* 493-511. https://doi.org/10.1177/0265532217710675

Omidian, T., Shahriari, H., & Siyanova-Chanturia, A. (2018). A Cross-disciplinary Investigation of Multi-Word Expressions in the Moves of Research Article Abstracts. *Journal of English for Academic Purposes, 36,* 1-14. https://doi.org/10.1016/j.jeap.2018.08.002

Ortega, L. (2003). Syntactic Complexity Measures and their Relationship to L2 Proficiency: A Research Synthesis of College-Level L2 Writing. *Applied Linguistics, 24,* 492-518. https://doi.org/10.1093/applin/24.4.492

Pallotti, G. (2015). A Simple View of Linguistic Complexity. *Second Language Research 31,* 117-134. https://doi.org/10.1177/0267658314536435

Paquot, M. (2019). The Phraseological Dimension in Interlanguage Complexity Research. *Second Language Research, 35,* 121-145. https://doi.org/10.1177/0267658317694221

Parkinson, J., & Musgrave, J. (2014). Development of Noun Phrase Complexity in the Writing of English for Academic Purposes Students. *Journal of English for Academic Purposes, 14,* 48-59. https://doi.org/10.1016/j.jeap.2013.12.001

Pho, P. D. (2008). Research Article Abstracts in Applied Linguistics and Educational Technology: A Study of Linguistic Realizations of Rhetorical Structure and Authorial Stance. *Discourse Studies, 10,* 231-250. https://doi.org/10.1177/1461445607087010

Ruan, Z. (2018). Structural Compression in Academic Writing: An English-Chinese Comparison Study of Complex Noun Phrases in Research Article Abstracts. *Journal of English for Academic Purposes, 36,* 37-47. https://doi.org/10.1016/j.jeap.2018.09.001

Song, R. M., & Wang, H. Y. (2019). Study on Syntactic Complexity of Chinese PhD Thesis Abstracts. *Journal of PLA University of Foreign Languages, 42,* 84-91.

Swales, J., & Feak, C. (2009). *Abstracts and the Writing of Abstracts.* The University of Michigan Press. https://doi.org/10.3998/mpub.309332

Taguchi, N., Crawford, W., & Wetzel, D. Z. (2013). What Linguistic Features Are Indicative of Writing Quality? A Case of Argumentative Essays in a College Composition Program. *TESOL Quarterly, 47,* 420-430. https://doi.org/10.1002/tesq.91

Wang, H. (2020). A Corpus-Driven Approach to Teaching Academic Writing for College Students. *Foreign Language Research, 1,* 49-55. https://doi.org/10.16263/j.cnki.23-1071/h.2020.01.007

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity.* University of Hawaii Press.

Yin, S., Gao, Y., & Lu, X. (2021). Syntactic Complexity of Research Article Part-genres: Differences between Emerging and Expert International Publication Writers. *System, 97,* Article ID: 102427. https://doi.org/10.1016/j.system.2020.102427