Chabahar Maritime University
Iranian Journal of English for Academic Purposes ISSN: 2476-3187
IJEAP, 2020, 9(1) (Previously Published under the Title: Maritime English Journal)

A Corpus-based Study of Noun Phrase Complexity in Applied Linguistics Research Article Abstracts in Two Contexts of Publication

[1]Mohammad Ahmadi
[2]Rajab Esfandiari*
[3]Abbas Ali Zarei

IJEAP- 2002-1503

Received: 2020-02-20 Accepted: 2020-03-24 Published: 2020-03-31

Abstract

Unlike conversation, academic writing is characterized by the frequent use of noun phrases which make it difficult for less proficient readers to process a text. Using a subset of Biber, Gray, and Poonpon's (2011) hypothesized developmental stages of writing, we analyzed noun phrase modifiers in applied linguistics research article (RA) abstracts between expert non-native English Persian writers and international writers. To that end, a 38,762-word corpus was constructed, consisting of 109 international academic research articles (RAs) and 100 Persian English-medium RAs randomly chosen from international peer-reviewed journals and Persian English-medium peer-reviewed journals. Using an automatic extraction computer program (PyCharm, version 3.4.), we tagged texts, identified noun phrase modifiers, and compared the normalized frequency of the modifiers between two writer groups. Independent-samples *t*-tests and chi-square tests of independence were run to analyze the data. The findings revealed that international writers differed significantly from Persian writers in the use of total noun phrase modifiers, relative clauses, and post-modifying prepositional phrases. Results from the analysis of lexical bundles indicated that Persian writers used lexical bundles to modify noun phrases more frequently than international writers. The findings of this study offer insights into the way expert international and non-native academic writers in applied linguistics make use of phrasal features for complexifying RA abstracts.

Keywords: RA Abstracts, Academic Writing, Noun Phrases, Phrasal Complexity

1. Introduction

Following Swales' (1990) seminal work which characterized rhetorical organization of research article introductions through moves and steps, many researchers adopted a genre-based approach to analyze academic writing (e.g., Cheng, 2019; Taylor & Goodall, 2019). In particular, rhetorical organization of abstracts has gained increasing attention over the past few years (e.g., Gholipour & Saeedi, 2019; Lores 2004; Samraj, 2005). Writing RA abstracts is undoubtedly a daunting task which requires considerable expertise to meet the expectations of discourse community in the field. In abstracts, writers demonstrate that their studies "have something worthwhile to say to gain the interest of the reader" (Hyland & Tse, 2005, p. 126). As Stotesbury (2003) put it, RA abstracts are evaluative in nature, since they provide the summary of the whole study which could be subsequently persuasive for the readers. Due to space saving requirements imposed by publishers in the era of information explosion, the need for compressed academic writing style seems more compelling than ever.

Recently, the study of RA abstracts has been on the rise as there has been a rise of experimental RAs in the new age of science (see Biber & Gray, 2010), because abstracts could be served as "stand-alone mini-texts" (Huckin, 2006, p. 93) which provide readers with the preview of

[1] PhD student of TEFL, ahmadi.m8362@gmail.com; Department of English Language, Faculty of Humanities, Imam Khomeini International University, Qazvin, Iran.

[2] Assistant Professor of TEFL, esfandiari@hum.ikiu.ac.ir; Department of English Language, Faculty of Humanities, Imam Khomeini International University, Qazvin, Iran.

[3] Associate Professor of TEFL, a.zarei@hum.ikiu.ac.ir; Department of English Language, Faculty of Humanities, Imam Khomeini International University, Qazvin, Iran.

the whole article. Through abstracts, readers are able to screen other sections of RAs like methods, results, and implications. Abstracts are ideal sites that allow writers to advertise and "sell" (Pho, 2008, p. 231) their academic works to readers who may become interested in reading the whole article.

Noun phrase modifiers play a key role in crafting RA abstracts, since compact discourse style could be achieved through noun phrase modification (Ruan, 2018). Academic writing relies heavily on noun phrase modifiers (Biber, Gray, & Poonpon, 2011) in order to use as much information as possible tightly packed into relatively few words (Halliday & Martin, 1993). In a series of studies, Biber, Conrad, Reppen, Byrd, & Helt (2002), Biber (2003), Biber, Gray, & Poonpon (2011), Biber (2014), Biber and Gray (2016) have challenged conventional measures of syntactic complexity, arguing that they are characteristic of spoken rather than written language. Their precise disciplinary corpus analyses revealed the stereotype that academic writing is more complex and explicit than conversation is not confirmed. Both conversation and academic writing are complex, but the way complexification is realized differs. In fact, conversational discourse relies heavily on short simple clauses, especially dependent clauses, while academic writing is dependent on noun phrases.

While much research has been done on the organizational patterns of RA abstracts, lexico-grammatical features and linguistic resources which are used to construct this section in contrastive analyses need further investigation. Accordingly, the current study is an attempt to investigate noun phrase complexity in applied linguistics RA abstracts in international and Persian English-medium journals. More specifically, this study intends to compare 15 noun phrase modifiers of attributive adjectives, relative clauses, nouns as pre-modifiers, possessive noun as pre-modifiers, of phrase (concrete/locative meanings), prepositions as noun post-modifiers other than of (concrete/locative meanings), -ed participle as post-modifiers, -ing participle as post-modifiers, attributive adjectives and nouns as pre-modifiers, of phrase (abstract meanings), prepositions as noun post-modifiers other than of (abstract meanings), preposition + nonfinite complement clause, complement clauses controlled by nouns, appositive noun phrases, and multiple prepositional phrases as post-modifiers with levels of embedding, between two corpora of international academic writers and Persian academic writers.

2. Literature Review

2.1. Noun Phrase Modifiers as Indices of Advanced Academic Writing

Recently, there has been a growing number of studies which indicate that complex noun phrases are hallmarks of advanced academic writing (e.g., Biber, et al., 2011, 2016; Kyle & Crossley, 2018; Norris & Ortega, 2009; Ravid & Berman, 2010; Ruan, 2018). In a recent study Biber and Gray (2016) reported that the use of compressed noun phrases has been increasingly favored in academic writing over the past 300 years. They further argued that the frequent use of subordinations is no longer the central feature of academic writing; rather, it is the property of conversational discourse. On the other hand, advanced academic writing is replete with the dense use of phrasal expressions.

The widespread assumption about academic writing is that it becomes more complex in terms of clausal embedding and subordination along the path of proficiency development (Casanave, 1994). Findings from research studies have shown that subordinate clauses are the distinctive features of academic writing (e.g., Hughes, 2005). However, the following excerpts which are taken from Staples, Egbert, Biber, & Gray (2016, pp. 2-3) question this widespread assumption.

(1) Selectivity [of the harvest [on Putauhinu Island]] TRANSLATES into large **differences** [in harvest rates [among weight **classes**]. [. . .] There is **evidence** [for such **links** [between **characteristics** [of young **individuals**] and life history **traits** [of adults]] [in many taxa]].

(2) Yeah, I'M HAVING fun. Well, yeah, they're probably GOING TO ASK me [to WRITE about it], [because I Think [I'm one of the first ones [that they've SENT out [to DO this]]]] so they'LL ASK me [to TELL them [how everything WORKED]] and I'M GOING TO SAY [it

WAS pretty amazing [how it all WORKED OUT]] (verbs are UNDERLINED and embedded clauses are marked in **[brackets]**)

Looking at excerpt 1 in which the main verbs are underlined, we can see that the whole clause only contains one main verb (translate) with no embedded clauses. Instead, complexity is realized in long phrasal expressions with a number of noun modifiers (pre-modifiers are in *italics*, post-modifiers are in [brackets], head nouns are **in bold type**). Excerpt 2, on the other hand, differs significantly from excerpt 1 which relies heavily on phrasal modification. The studies carried out by various researchers have confirmed that grammatical structures in academic writing normally involve the features similar to excerpt 1 (e.g., Biber & Gray, 2016; Lu, 2011). Biber & Gray (2016) documented that the clausal features are more frequently found in written register than in academic writing. This highlights the importance of phrasal features in academic prose, especially for student writers (Parkinson & Musgrave, 2014).

Building on the differences between spoken and written register with regard to predominant complexity features, Biber, et al. (2011) hypothesized developmental stages in academic writing based on the results obtained from a large corpus study. The stages are based on syntactic functions of finite dependent clauses, nonfinite dependent clauses, and dependent phrases. As Staples, et al. (2016) noted, "these stages are based on the premise that novice academic writers start with clausal complexity features most common in speech, and then gradually develop proficiency in the dense use of phrasal complexity features associated with specialist academic writing" (p. 5). Fifteen phrasal features could be extracted from Biber, et al.'s (2011) hypothesized developmental stages (see Table 1 for more information).

Table 1: Biber, et al.'s (2011) Hypothesized Noun Phrase Developmental Stages

Stage	Grammatical Structure	Examples from our corpus
2	Attributive adjectives	Significant result
	Relative clauses	Explanation that comes to mind
3	Nouns as pre-modifiers	System complexity
	Possessive noun as pre-modifiers	Speaker's speech
	Of phrase (concrete/locative meanings)	Community of scholars
	Prepositions as noun post-modifiers other than of (concrete/locative meanings)	Members in the research field
4	-ed participle as post-modifiers	Context related to their professional experience
	-ing participle as post-modifiers	The tasks facilitating communication
	Attributive adjectives, nouns as pre-modifiers	Adult language learner
	Of phrase (abstract meanings)	Realization of thanking
	Prepositions as noun post-modifiers other than of (abstract meanings)	Development for writers and speakers
5	Preposition + nonfinite complement clause	Methods for classifying texts
	Complement clauses controlled by nouns	Misconceptions that have arisen regarding student writing
	Appositive noun phrases	Multidimensional (MD) analysis
	Multiple prepositional phrases as post-modifiers, with levels of embedding	Product of the interaction of grammar with the context of production

As Biber, et al. (2011) put it, the developmental index shown in table 1 starts from one of the intermediate stages of noun phrases functioning as constituents in other clauses to the last stage of dense use of phrasal (non-clausal) dependent structures that function as constituents in noun phrases. While the intermediate stages are mainly characterized by noun pre-modifiers, the final stages represent dense use of phrasal features as noun post-modifiers which are understood to approximate advanced academic writing.

Studies on noun phrase complexity in L1 and L2 is scanty. More specifically, few studies have investigated noun phrase complexity by considering L1 background as a potential source influencing the type and frequency of noun modifiers in RAs. In one of the few studies, Lan and Sun (2019) explored noun phrase complexity in L2 writings of Chinese first-year college students. Through comparing the frequency of noun phrase modifiers in L2 students' writings and academic journal articles, they found that there were significant differences between the writings of the two groups in terms of both total frequency and types of noun modifiers. However, their study drew on the findings obtained from two incomparable corpora of RAs and students' writings. As Swales (1990) put it, RAs constitute a distinctive type of genre because most of them undergo the process of peer-reviewing for publication which in turn pushes them into special kind of academic prose approved by the journals. On the other hand, students' writings belong to Swales' (1996) category of occluded genre (Nesi & Gardner, 2012) because as Loudermilk (2007, as cited in Nesi & Gardner, 2012) point outed, "students rarely show their coursework to anyone other than their tutors" (p. 28).

Additionally, many features of phrasal modifiers lend themselves to multi-word expressions (e.g., attributive adjective + noun = *significant effect*; adjective-adjective compounds = *a cross-sectional investigation*; post-modifier proposition = *implication for pedagogy and future research*). As Myles (2012) stated, studies of syntactic complexity need to take into account the role of pre-constructed multi-word units and formulaic expressions explicitly learned as part of vocabulary training. In Myles' words, "formulaic sequences are very common in early L2 productions and enable learners to communicate in spite of limited linguistic means so that they appear to be more advanced in the L2 than they actually are" (p.71). A special type of formulaic sequences is lexical bundles. Biber and Conrad (1999) characterized them as "simply sequences of word forms that commonly go together in natural discourse" (p. 990). The way published international and nonnative English expert writers construct noun phrases using lexical bundles may shed new light on the construct of syntactic complexity in academic writing.

2.2. Recurrent Word Combinations in Academic Writing

In literature, different terminologies have been used to address multi-word sequences. 'Recurrent word combinations' is an umbrella term which refers to a sequence, continuous or discontinuous, of words that is, or appears to be prefabricated; that is, it is stored and retrieved as whole rather than individually (Wray, 2002). One of the prominent features of recurrent word combinations is their abundance in academic register as nearly 21 percent of the words in Biber, Johansson, Leech, Conrad, and Finegan's (1999) corpus of academic prose consisted of recurrent word combinations, for which they used the phrase lexical bundles.

The study of recurrent word combinations is gaining increasing attention as a large number of studies have investigated word co-occurrences since 1987 when Sinclair proposed 'idiom principle' in an attempt to demonstrate the formulaicity of the language. Within academic writing, recurrent word combinations play a leading role, because "different registers and genres often carry their own unique vocabulary, forms of expression, and conventionalized word combinations that are required for acceptance into the target community" (Appel & Wood, 2016, p. 55). It is impossible to imagine how dull a piece of academic prose might look like if it were devoid of these sequences. As Wray and Perkins (2000) stated, multi-word expressions serve the purpose of shortcuts by allowing language users to process and retrieve the sequences as a whole rather than individually on each occasion.

When the measures for identifying multi-word expressions become purely quantitative, we may think of the phrase lexical bundles. The term was first coined by Biber, et. al. (1999), who set quantitative criteria such as frequency and range for identification of these sequences. Lexical bundles differ from idioms in that there is a transparency of meaning, that is, the meaning of whole could usually be identified by adding up the meaning of each individual component. According to Biber and Barbieri (2007), lexical bundles are not structurally complete, but they perform important discourse functions.

Drawing on the findings of the previous studies on phrasal complexity, the current study is a response to the call by Biber, et. al. (2011) for more empirical investigation of noun phrase modifiers in academic writing. The current study is an attempt to analyze noun phrase complexity in applied linguistics RA abstracts in international and Persian English-medium journals. Accordingly, this study is guided by the following two research questions:

Research Question One: Are there any significant differences in the frequency of 15 noun phrase modifiers between international and Persian expert writers?
Research Question Two: To what extent does phrasal complexity in international and Persian English-medium journals rely on formulaic patterns?

3. Methodology

3.1. Construction of the Corpus

Two corpora were developed to carry out the quantitative analyses in this study: International Corpus (IC) and Persian Corpus (PC). They include applied linguistics RA abstracts in international and Persian English-medium peer-reviewed journals. RA abstracts were chosen because of three main reasons: (a) They have manageable length, which makes cross comparisons easier than other sections, (b) they are "a compact genre" (Jiang & Hyland, 2017, p.3) which, the present researchers believe, would be suitable for investigating noun phrase modifiers which are mainly used for compressing the text, and (c) RA abstracts are getting increasing attention since they "foreground important claims, minimize methodology and background statements, and pack information into visuals" (Hyland, 2000, p.86).

The IC comprised the RAs written by international writers. In the present study, international writers refer to native English writers from English-speaking countries and non-native English writers from non-English speaking writers. The journals comprising RAs in IC are diverse in their focus. To ensure the credibility of the journals, two criteria were set: year of publication and H index. The former is selected because the history of established journals generally contributes to the journal's scientific background in the field. Older journals "have had time to build credibility, in contrast to new journals with only a handful of issues" (Hutter, 2015, p. 26). H index also represents the journals' credibility by combining "publication activity and citation influence" (Öchsner, 2013, p. 51). As Harzing and van der Wal (2008) put it, H index has at least two advantages over traditional Thomson ISI journal impact factor (JIF). First, H index is not influenced by highly cited articles because it is not based on mean scores. Second, it is not impacted by artificially fixed time horizon. International English-medium journals had to meet the criteria of the minimum publication history of 30 years and H index of 25. Accordingly, five International English-medium journals, as shown in Table 3, were selected to be included in the IC.

The PC included English-medium peer-reviewed journals published in Iran. To choose Iranian journals, the present researchers could not set the criterion of H factor, since most of them are not indexed in Web of Science, making it impossible to compare them with indexed journals. Nevertheless, the selected journals satisfied the criterion of approval granted by the Iranian Ministry of Science, Research and Technology, so all of them (see Table 3) are research-based and follow the strict procedures for publication of manuscripts as set by the Ministry. Accordingly, in order to have comparable number of journals in two corpora, five journals were also selected for inclusion in PC based on their years of publication. The articles were selected randomly from IC and PC through an online randomizer program (RNG).

Table 2: Descriptive Information of the Corpora of Research Article Abstracts

	No. of texts	Mean number of words	Total number of words
International Corpus	109	177.68	19368
Persian Corpus	100	193.94	19394

Table 3: Titles of Journals and Criteria for their Selection

Journal	Years of Publication	H factor
Language Learning	1948-1953, 1955-1956, 1958-ongoing	38
TESOL Quarterly	1981-ongoing	36
Modern Language Journal	1916-1996, 1998-2001, 2005-ongoing	36
Journal of Pragmatics	1977-ongoing	35
English for Specific Purposes	1980-1981, 1986-ongoing	25
Iranian Journal of Applied Language Studies	2009-ongiong	—
Journal of Teaching Language Skills	2009-ongiong	—
Journal of English Language Teaching and Learning	2010-ongiong	—
Journal of Language and Translation	2010-ongiong	—
Journal of Research in Applied Linguistics	2010-ongiong	—

All RAs followed IMRD format and were published between 2015 and 2018. The collection of recently published RAs characterizes "the present day" trends in academic writing (Biber & Gray, 2016). We exercised great care to make both corpora comparable in length. As Crawford and Csomay (2016) pointed out, corpus balance is understood to be one of the critical features of corpus building. Corpus balance could be defined in terms of number of texts and number of words. However, Crawford and Csomay (2016 as cited in Ansarifar, Shahriari, & Pishghadam, 2017) claimed that "frequency comparisons are done on the basis of the number of words, not by the number of texts" (p. 62). Therefore, our corpora included almost the same number of words, as shown in table 2.

The journals chosen in each corpus were different with regard to number of issues and number of articles published in each issue (see Table 2). Accordingly, journal articles were randomly selected from each corpus. Persian journals contained greater number of words in the abstract section (mean = 193.94) than International journals did (mean = 177.68). However, this problem was carefully addressed by normalizing the frequency count of grammatical features of interest in 1,000 words (see Biber & Barbieri, 2007) in order to allow for comparability of the data.

3.2. Grammatical Features of Interest

The present study aimed to investigate 15 noun phrase modifiers as identified by Biber, et al. (2011). Noun phrase modification features presented in Table 1 are obtained from developmental stages of syntactic complexity proposed by Biber, et al. (2011). The developmental index entails five stages which are categorized based on three grammatical types: Finite dependent clauses, nonfinite dependent clauses, and dependent phrases. In this study, the purpose was to examine (1) finite dependent clauses including relative clauses as noun modifiers, complement clauses controlled by nouns; (2) nonfinite dependent clauses including, -ing and -ed participles as noun post-modifiers, and preposition + nonfinite complement clauses as post-modifiers; and (3) dependent phrases including, attributive adjectives, participles, nouns as pre-modifiers, possessive nouns, of phrases as noun post-modifiers, other prepositional phrases as noun post-modifiers, adjectives, noun as pre-modifiers, appositives, and multiples prepositional phrases as noun post-modifiers.

3.3. Identification of Noun Phrase Modifiers

Noun phrase in this study was operationalized as "a string of words with a lexical noun as its head" (Ravid & Berman, 2010, p. 6). That is, those structures that contained a determiner and head noun or simply a noun were considered simple noun phrases. A noun phrase in its basic form contains an optional determiner and a head noun and any additions to these patterns result in complex grammar (Biber, et al., 2011). Since this study investigated noun phrase complexity in academic writing, simple noun phrases were not included for further investigation. Prepositional phrases received special treatment in this study, because they can either function as noun post-modifiers or as

adverbials. Those functioning as adverbials were not included for analysis because adverbials modify the preceding verb rather than the preceding noun.

Noun phrase modifiers included in this study were coded according to Biber, et al.'s (2011) hypothesized developmental stages of noun phrase complexity, e.g., attributive adjectives (s*ignificant result*), nouns as pre-modifiers (*explanation that comes to mind*), possessive nouns (*speaker's speech*), of phrases as noun post-modifiers (*community of scholars*), other prepositional phrases as noun post-modifiers (*members in the research field*), -ed participle as post-modifiers (*context related to their professional experience*), appositives (*Multidimensional (MD) analysis*) , and multiples prepositional phrases as noun post-modifiers (*product of the interaction of grammar with the context of production*).

Automatic text analysis tools for assigning grammatical tagging, syntactic dependencies, and constituency parsing are extensively used in corpus-based studies (e.g., Lu, 2011; Lu & Ai, 2015; Staples, et al., 2016). This is because manual coding of data is a labor-intensive process (especially with large corpora) which requires a great deal of expertise on part of the coders. However, they have limited success in distinguishing prepositional phrases functioning as post-modifiers from those as adverbials (Biber & Gray, 2016, p. 65). Moreover, "number of noun phrase modifiers could not be automatically done by corpus tools" (Ruan, 2018, p.7). Accordingly, once the quantitative analysis was done, the researchers of the present study conducted the qualitative check of previously identified lexico-grammatical features by the program.

The first phase of noun phrase identification included automatic extraction of noun phrase modifiers. Initially, the texts in each corpus were automatically tagged through a computer program called Stanford Core NLP Version 3.9.2. Stanford core NLP is a free tool that assigns part-of-speech to the words as well as their syntactic complexity dependencies. It provides a set of human language technology tools (Manning, et al., 2014). Depending on the datasets on which the analysis is carried out, the accuracy of Stanford Core NLP is reported to be between 97.21 and 97.67 (Manning, 2015). Part-of-speech (POS) tagging is a preliminary step for more advanced processes needed for syntactic complexity analysis (i.e., constituency and dependency parsing) and provides some of the information needed for fine-grained syntactic complexity analyses (Song & Chambers, 2014). Then, using a special Pycharm program, which is run on Python environment, noun modifiers were extracted. The following is an example of noun phrase modifiers which were extracted by means of the program: *The imaginary situation evoked by task was also found to bring about different means of learner involvement.*

Table 4: Noun Phrase Modifiers Identified by Automatic Extraction Tool

Noun Phrase	**Count**	**Text**
Noun + noun	1	1- learner involvement
Adjective + noun	2	1- imaginary situation
Adjective + noun	2	2- different means
Noun + past participle	1	1- situation evoked
Noun + *of* preposition	1	1- means of

As can be seen, the text has been tokenized into sentences and POS tagging was done individually for each sentence. The total number of lexico-grammatical features of interest was also counted (table 4). This allowed the coders to check the accuracy of the program with regard to POS tagging.

The qualitative phase started with a discussion session in which the coders, who were trained in applied linguistics and had an extensive experience in syntactic coding, shared their ideas in order to arrive at an accurate understanding of the coding scheme used in this study. Then, the coders tagged 10 percent of the corpus in order to calculate the accuracy of the program. Total accuracy rate of the program was around %95 with the highest rate for attributive adjectives and the lowest rate for prepositional phrases. Then, the

Coders started the manual check of the extracted features of the RA abstracts. In terms of disagreement between the coders, a third experienced coder in corpus linguistics, who was also

trained in applied linguistics, coded the data. Then, the agreement between the coders reached as close as 100%.

3.4. Identification of Lexical Bundles

The first step in identifying lexical bundles was deciding on the length of word sequences. It was an important decision because longer sequences could drop those modifying features comprising shorter number of words. For example, the bundles like *English language proficiency* would not be identified if the sequence of four words was set as the criterion of length for identification. The decision was based on the fact that the length of noun phrase modifiers rarely exceeds three words. Accordingly, it was decided that 3-word bundles could better fulfill the purposes of the current study.

The next criterion concerns cut-off frequency, which is set in the studies adopting a frequency-based approach for identifying lexical bundles. That determines the number of times a particular set of words occurs in a particular corpus. However, "the actual frequency cut-off used to identify lexical bundles is somewhat arbitrary" (Biber, et al., 2004, p.376). Previous studies have used different cut-off points for the identification of word clusters, which range from as low as 10 times per million words (Biber, et al., 1999), set to 20 times per million words (Hyland, 2008b), and as high as 40 per million words (Biber, Conrad, & Cortes, 2004). In this study, we steered a middle ground and set the cut-off frequency at 20 times per 1 million words.

Range was the third criterion needed to be identified in the present study to guard against writers' idiosyncrasies. Unlike frequency, range has to do with the number of texts a lexical bundle needs to recur. Previous studies have used between three and five texts (Chen & Baker, 2010; Biber, et al., 2004) depending on corpus size. The corpora in the present stud were small in size, so for a word combination to qualify as a lexical bundle, it had to occur in at least three texts.

3.5. Statistical Analyses

In order to answer the first research question, grammatical features of interest were identified through a procedure discussed in section 3.3. Then, they were normalized to 1000 words to ensure the comparability of the features between the two corpora. Independent-samples t-tests were used in order to compare the differences between international and Persian writers in terms of frequency of use of 15 noun phrase modifiers. Since multiple comparisons were performed, Bonferroni post-hoc adjustment was used to adjust the alpha level which was set at $p < 0.002$ after correction.

In order to answer the second research question, we used AntConc version 3.4.4.0 (See Anthony, 2019) to identify lexical bundles. AntConc uses the plain text and identifies and sorts clusters of specified size based on the given criteria (e.g., minimum frequency and minimum range). Then, the identified bundles were divided by the number of modifiers in each stage. Then numbers obtained in each group was compared against those of the other group by running Chi-square test for independence. Since multiple comparisons are made (5 comparisons), Bonferroni adjustment is applied and the new alpha level of $p < 0.01$ was set after correction.

4. Results

4.1. Results

4.1.1. Investigation of the First Research Question

Initially, the normalized frequency of noun phrase modifiers in each corpus is presented (Table 5). Then, the present researchers followed a more detailed investigation of the distributional pattern of noun phrase modifiers in each corpus in order to unveil the differences that might exist between Persian and international academic RA abstracts with regard to noun phrase modifiers.

According to Table 5, international writers used more modifiers on average (Mean = 267.13) than Persian writers (Mean = 246.77) and the difference was statistically significant at $p < 0.002$. The effect size of 0.500 which was calculated by Cohen's d indicated a medium effect of independent variable (Cohen, 1988). For more detailed analysis of types of noun phrase modifiers,

we divided the total modifiers into two groups of pre- and post-modifiers. As shown in Table 5, the international writers used more pre- and post-modifiers than Persian writers did. However, only the difference between post-modifiers between the two writer groups was statistically significant ($p < 0.002$).

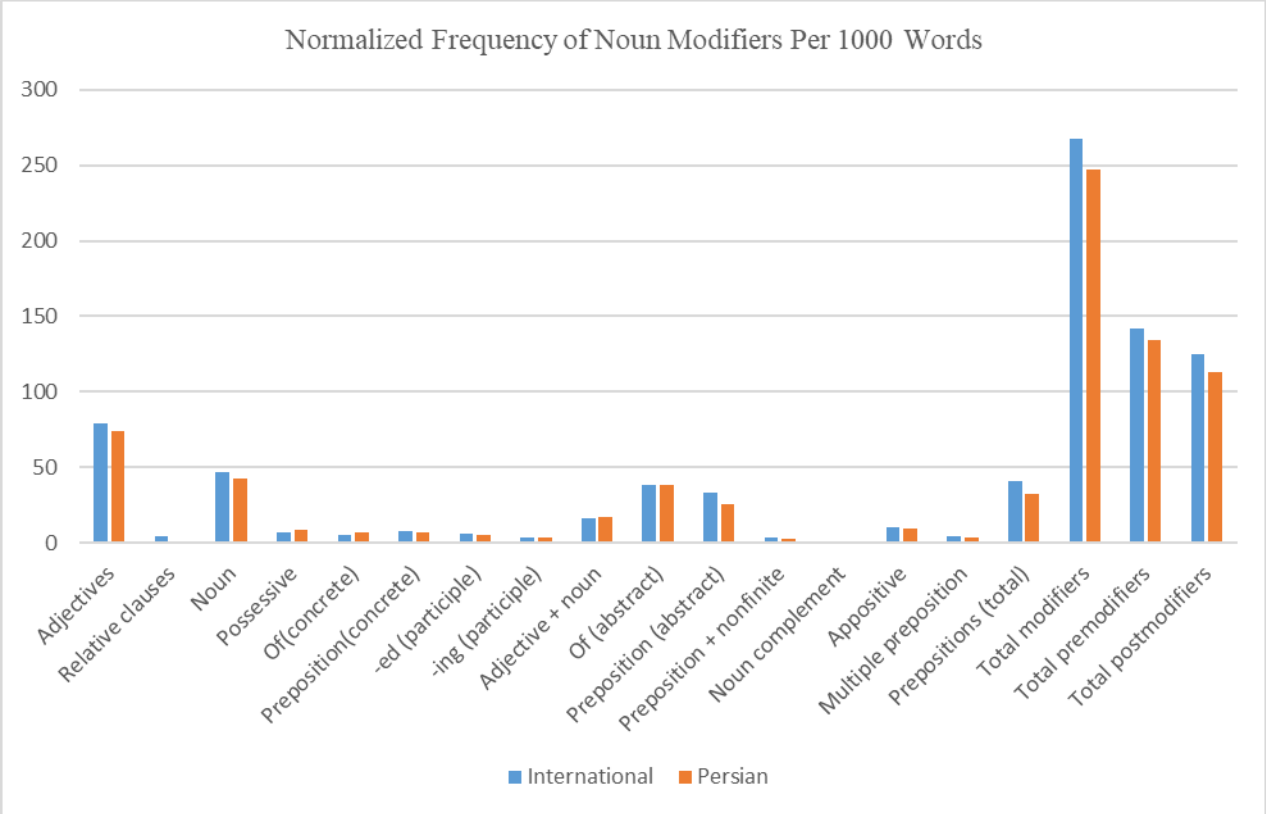

Figure 1. Distribution of noun phrase modifiers in two groups of writers

Normalized frequency of 15 noun phrase modifiers was compared between international and Persian academic writers. Since multiple comparisons were made, Bonferroni adjustment to the alpha level was made and the new alpha level of 0.002 was set. The results obtained from multiple comparisons by running independent-samples t-test revealed that there were significant differences ($P < 0.002$) in the mean scores of only two of the noun phrase modifiers of "relative clause" and "post modifying prepositions". The effect size calculated by Cohen's d also confirmed the difference between the two groups. The value of effect size was 0.711, and 0.500 for "relative clauses", and "post modifying prepositions" respectively meaning medium "magnitude of the difference between the groups" (Pallant, 2013, p. 250) according to Cohen's (1988) classification of effect size.

Table 5 Comparison of Occurrence of Noun Phrase Modifiers between the Persian Corpus and International Corpus

Modifiers Type	Group	Normalized Mean	SD	*P*(*t*-test)	Cohen's d
Adjectives	International	79.05	17.32	0.818	0.216
	Persian	74.14	16.67		
Relative clauses	International	4.07	5.55	0.000*	0.711
	Persian	1.04	2.34		
Noun	International	46.65	24.62	0.271	0.153
	Persian	43.07	22.12		
Possessive	International	6.96	8.99	0.226	0.168
	Persian	8.43	8.47		

Of (concrete)	International	5.5	7.74	0.222	0.17
	Persian	6.97	9.55		
Prepositions (concrete)	International	8.22	8.94	0.228	0.167
	Persian	6.69	9.35		
-ed participle	International	5.83	6.65	0.617	0.069
	Persian	5.36	6.98		
-ing participle	International	3.61	4.57	0.744	0.045
	Persian	3.82	4.8		
Adjective + Noun	International	16.13	11.47	0.66	0.06
	Persian	16.89	13.67		
Of (abstract)	International	38.43	16.88	0.918	0.014
	Persian	38.66	15.66		
Prepositions (abstract)	International	32.93	16.5	0.001*	0.500
	Persian	25.78	13.53		
Prepositions + nonfinite	International	3.83	5.67	0.036	0.292
	Persian	2.37	4.13		
Noun + complement	International	0.76	2.25	0.025	0.301
	Persian	0.21	1.07		
Appositives	International	10.49	8.89	0.6	0.072
	Persian	9.82	9.71		
Multiple prepositions	International	4.47	5.18	0.167	0.192
	Persian	3.51	4.76		
Prepositions (total)	International	41.14	18.37	0.000*	0.500
	Persian	32.46	16.39		
Total modifiers	International	267.13	42.26	0.000*	0.500
	Persian	246.77	40.61		
Total pre-modifiers	International	141.83	30.18	0.056	1.152
	Persian	134.101	27.84		
Total post-modifiers	International	125.29	28.05	0.002*	1.182
	Persian	112.67	28.67		

*Note. The values are significant at $p < 0.002$.

In addition, Figure 1 shows that attributive adjectives are the most frequently used types of noun modifiers in both groups. On the other hand, multiple prepositional phrases were the least frequent types of modifiers in two groups. With the exception of five modifiers of "possessives", "of (concrete)", "-ing participle", "adjective + noun", and "of (abstract)", the writers in international group used more modifiers than the Persian writers. The most striking difference was between "total modifiers" followed by "total post modifiers", and "prepositions total".

4.1.2. Investigation of the Second Research Question

The second research question was concerned with determining the extent to which noun phrase modifiers depended on lexical bundles. Once 3-word sequences were automatically identified through the concordance tool, they were manually checked for further analyses. The extracted bundles were checked against the problem of overlap, or subsumption, which could distort the results obtained (Chen & Baker, 2010). They could happen when two or more 3-word bundles are actually parts of a longer bundle. For example, the bundles "significant differences between", and "differences between the" occurred frequently in our corpus, but these bundles overlap because they are both subsumed under the bundle "significant differences between the". In this case, the more frequent bundle was kept and the less frequent one was discarded. The extracted bundles were

categorized according to the functions they served as noun phrase modifiers in the corpus. Accordingly, five groups of lexical bundles were identified which are presented in Table 6.

Table 6: Lexical bundles as noun modifiers in L1 Persian and international corpora

Grammatical structure	International	Persian	$\chi 2$
Of phrases as post-modifiers	152 (18%)	199 (22.3%)	0.000*
Attributive adjectives	68 (4.5%)	102 (7.2%)	0.002*
Attributive adjectives and Nouns	16 (5.2%)	42 (13%)	0.001*
Nouns as pre-modifiers	10 (1.1%)	26 (3.2%)	0.003*
PPs other than *of*	46 (5.8%)	40 (6.4%)	0.618
Total	**292 (6.7%)**	**409(10%)**	**0.000***

*Note. The values are significant at $p < 0.01$.

As shown in Table 6, identified bundles include *of* phrases as post modifiers, attributive adjectives, attributive adjectives and nouns as pre-modifiers, nouns as pre-modifiers, and prepositional phrases other than *of* as post-modifiers. *Of* phrases as post modifiers, and nouns as pre-modifiers are the most and the least frequent clusters in the two corpora respectively.

According to Table 6, Persian writers used more lexical bundles as noun modifiers of all types with the exception of PPs other than of compared to international writers, and the results obtained by running Chi square test of independence revealed that the difference between the two groups are statistically significant in all features with the exception of PPs other than *of*. In addition, the difference between the proportion of total lexical bundles to total number of noun phrase modifiers in two groups of writers reached statistical significance ($p < 0.01$). Altogether, the writers in international corpus used 65 different bundles while the writers in Persian corpus used 61 different bundles. To be more precise, only about 28 percent of the bundles were shared between international and Persian writers. Other bundles were either nonexistent in IC, or were used with different word combinations, and sometimes with different syntactic functions. For example, almost half of the occurrences of the bundle "the use of" in PC was part of a larger bundle "variation in the use of"; however, in IC in more than half of the cases, the bundle was used immediately after a verb to function as the direct object.

Table 7: The Top 3-word Lexical Bundles by International and Persian Writers

International	Normalized Mean	Persian	Normalized Mean
a foreign language	87.77	the present study	175.31
the use of	56.79	the results of	159.84
the development of	56.79	Iranian EFL learners	108.43
in second language	46.47	the use of	108.28
the relationship between	46.47	the current study	67.03
the role of	46.47	two types of	51.56
the effectiveness of	41.31	findings of the	41.25
the present study	41.31	the effect of	41.25
analysis of the	36.14	difference between the	41.25
in applied linguistics	36.14	a number of	36.09
the impact of	30.98	the impact of	36.14
the context of	30.98	different levels of	36.09
of this study	30.98	implications of the	36.09
the aim of	30.98	the relationship between	36.09
the analysis of	30.98	the role of	36.09
a lack of	30.98	in second language	30.94
of second language	25.82	the analysis of	30.94
reports on a	25.82	the control group	30.94
a set of	25.82	the process of	30.94

the need for	25.82	a foreign language	25.78
a case study	20.65	learners in the	25.78
a corpus of	20.65	part in the	25.78
a range of	20.65	the effects of	25.78
differences between the	20.65	the participants of	25.78
research on the	20.65	in applied linguistics	25.78
second language learning	20.65	in language learning	25.78
the implications of	20.65	the development of	25.78
the value of	20.65	the field of	25.78
in English-medium instruction	20.65	the importance of	25.78
pedagogical implications for	20.65	a group of	20.62
the field of	20.65	data analysis showed	20.62
a number of	15.49	Iranian EFL teachers	20.62
in the classroom	15.49	a series of	20.62
a series of	15.49	the nature of	20.62
important role in	15.49	the perspective of	20.62
analyses of the	15.49	the quality of	20.62
applications of the	15.49	the translators of	20.62
part of a	15.49	a mixed method	15.47
Total	**1461.17**	**Total**	**1974.83**

Table 7 provides a more detailed picture of the bundles used by two groups of writers. As can be seen, in PC the first four bundles are very common occurring more than 100 times per million words with the most common of them 175 times per million words; however, in IC, the difference between the most frequent bundles and the less frequent ones is less extreme. Overall, Persian writers used more lexical bundles (1974.83) than international writers (1461.17).

5. Discussion

The first finding of our study was that international writers used significantly more noun phrase modifiers than Persian writers did. By dividing the noun phrase features into two groups of pre- and post-modifiers, we found that international writers used post-modifiers more frequently than did Persian writers. This is in line with some of the findings of previous studies that have revealed that more proficient writers rely more heavily on post-modifiers than less proficient writers do (Ansarifar, et. al., 2017; Parkinson & Musgrave, 2014; Ruan, 2018), because post-modifiers are distinctive features of more advanced stages of academic writing, as shown in Biber, et. al.'s (2011) model.

The widespread assumption that academic writing is more explicit than conversation is no longer attested, as shown in this study and other similar studies (Gardner, Nesi, & Biber, 2018; Staples et al., 2016); rather, it is now confirmed that frequent use of noun phrase modifiers and lack of explicit relations between them make the text less explicit. That said, a trade-off between the economy of space and clarity of expression in academic writing is usually made (See Biber & Gray, 2016). However, expert academic writers alleviate the tension between lack of explicitness in meaning and frequent use of noun phrases by using post-modifying prepositional phrases (Biber & Gray, 2010). As Wu, Mauranen, and Lei (2020) argued pre-modifiers like *corn oil* are less explicit than post-modifiers like *oil from corn.* Excerpt 3 shows how international academic writers used post-modifying prepositional phrases to make the text more explicit. "With a preposition in between, the internal logical relations of complex nominals become explicit, thus mitigating the burden of meaning processing" (Wu, et. al., 2020, p. 9).

(3) The marketization **of higher education in the UK and elsewhere** has attracted a great deal **of attention (and criticism) from applied linguists in recent years**, but

there is still little linguistic evidence **of its impact on the actual value system of academic institutions**.

The next finding of our study was that attributive adjectives were the most frequent types of noun phrase modifiers in both corpora. Biber, Gray, and Poonpon (2016) reported that unlike science writing, humanities employ dense use of attributive adjectives. The frequent use of attributive adjectives in academic writing has also been reported in previous studies (e.g., Lan & Sun, 2019; Parkinson & Musgrave, 2014; Ruan, 2018). Attributive adjectives are of particular importance in academic writing because they are not only valid tools for compressing the text, but also have a number of semantic functions. Biber et al. (1999) argued that attributive adjectives perform at least three functions of describing: categories of size, evaluation, and classifiers, but the most striking pattern is the reliance on classifiers (e.g., pragmatic competence).

We also found that prepositional phrases were the most common type of post-modifiers in both writer groups. This is not surprising because post-nominal modifications, especially of genitive expressions, are very common in academic writing (Cortes, 2004), because they cover a wide variety of meanings and functions to "elaborate logical (particularly temporal) or textual connections between elements of an argument" (Hyland, 2008a, p. 52). As Biber, et. al. (2011) put it, prepositional phrases occur fifteen times as frequently in academic writing as relative clauses.

A surprising finding of our study was that relative clauses occurred four times more frequently in IC than in PC. The normalized frequency of occurrence of relative clauses is 4.07 in international corpus and 1.04 in Persian corpus. These normed values are lower than that of 7 per 1,000 words in Biber, et. al.'s (2011) study. However, our finding confirms that of Ansarifar, et. al. (2017), who reported the normalized frequency of 4.92 per 1,000 words in the corpus of native English expert academic writing. Previous studies have distinguished four types of relative clauses based on the functions the head noun and relative pronoun fulfill in the sentence, which include object subject (OS), object object (OO), subject subject (SS), and subject object (SO) (see appendix A for some sentence examples of these relative clauses). The first represents the function of the head noun in the main clause and the second shows the function of the relative pronoun in the relative clause. A close examination of the relative clauses used by Persian writers in our study revealed that more than 86 percent of the occurrences followed OS order. In fact, only three occurrences followed other orders (only OO order). Structural differences between the Persian language and the English language may explain our unexpected finding. The Persian language is a verb-final language with subject-object-verb word order; by contrast, the English language follows a subject-verb-object order. This difference could be best explained by Hamilton's (1994) SO Hierarchy Hypothesis (SOHH) which posits that the difficulty of four types of relative clauses depends upon the function of head noun, and the function of relative pronoun which determines processing discontinuity. Processing discontinuity is the result of (a) a relative clause which interrupts the processing of information, (b) the distance between relative pronoun and its trace within the relative clause (see appendix B for examples of discontinuity).

As Marefat and Rahmani (2009) put it, the Persian language does not allow right branching of relative clauses. As a result, all of the RCs are center embedded. Therefore, OS would be the easiest structure for L1 Persian language writers to use and underuse the other three relativized types. The fourth finding of our study was that post-modifying prepositions (with abstract meaning) occurred more frequently in IC than in PC, and the result was statistically significantly different. Greater reliance on post-modifying prepositional phrases is the hallmark of advanced academic writing (Jiang, Bi, & Liu, 2019; Taguchi, Crawford, & Wetzel, 2013). Using post-modifying prepositional phrases by advanced academic writers is one of the sound strategies to "create dense information structure with few words" (Biber & Gray, 2016, p.191) and to establish explicit relations between the modifiers. Compressed structures are phrasal and include noun pre-modifiers and prepositional phrases, both of which lack verbs (Biber & Gray, 2010). Persian writers' less frequent use of post-modifying prepositional compared to international writers implies that Persian writers may not be aware of the valuable functions of these post-modifiers or may have not mastered less frequent prepositional phrases (other than *of*), probably due to insufficient exposure.

The results obtained corroborate those of Parkinson and Musgrave (2014) and those of Taguchi, et. al., (2013), where more proficient academic writers used more instances of post-modifying prepositional phrases compared to less proficient ones.

The findings of the second phase of our study revealed that the total number of lexical bundles used in PC was significantly more frequent than the one used in IC. That simply means that Persian writers were more reliant on lexical bundles in complexifying their academic writing. The more frequent use of lexical bundles by Persian academic writers compared to their international counterparts could be accounted for by the "more formulaic nature" of Persian academic writers' language resources and the fact that Persian writers needed to adopt "a more conciliatory approach" to constructing academic abstracts (Hyland, 2008a, p. 50). The results of the current study conform to those of Hyland (2008a), Liu and Liu (2009), and Wei (2007). Hyland, for example, documented that MA students employed more clusters than doctoral-level students who, in turn, used more clusters than professional academic writers. However, he argued that in addition to different language resources that the three groups relied on, the effect of genre on the writers' performance also had to be taken into consideration, since MA theses, which belong to a pedagogical genre, is different from those of PhD dissertations and RAs. The results of our study, though, ran counter to those of Chen and Baker (2010) who unveiled that published academic writing drew on a wide range of lexical bundles while student writing employed the smallest range. Persian writers relied more heavily on bundles probably because they have processing advantage and "they are not subject to generation or analysis by the language grammar" (Wray & Perkins, 2000, p.1). However, as Pawley and Syder (1983) stated, "despite the apparent ease with which they are adopted during learning, it is often the failure to use native-like formulaic sequences that ultimately marks out the advanced L2 learner as non-native" (As cited in Wray & Perkins, 2000, p.2).

One interesting finding of our study was that although Persian academic writers used noun-modifying lexical bundles more frequently than international writers, a great number of the bundles that were used by Persian writers were not found in the IC or were used far less frequently. As shown in table 7, bundles like "the use of" occurred approximately twice more frequently in IC than in PC. A relatively frequent bundle like "two types of" in PC was not found in IC, as Extract 5 shows. A relatively frequent bundle like "the effectiveness of" in IC did not recur in PC, as shown in Extract 6. Both writer groups, however, seemed to grasp the importance of post-modifying prepositional phrases, because noun phrase + *of* was the most frequent bundle in both corpora. *Of*-prepositional phrase is favored in academic writing (Biber & Gray, 2016), because of its major role in compressing the text as well as its potential to be used recurrently.

> (5) The occurrences of the **two types of** nominal expressions were counted and normalized. The **two types of** integrated tasks produced features that shared to a large extent.
>
> (6) The purpose of the study was to investigate **the effectiveness of** a vocabulary task in terms of its impact on vocabulary acquisition. For the learners in our study, **the effectiveness of** feedback depended on other factors.

6. Conclusions and implications

The findings of our study revealed that international academic writers used significantly more post-modifiers on average than Persian academic writers. On this basis, we conclude that international RAs are more complex than Persian English-medium RAs and that Persian academic writers may not have fine-tuned their syntactic repertoire to phrasal features of academic writing. We also found that Persian writers relied more heavily on lexical bundles for constructing noun phrase modifiers than international writers did. Linguistic knowledge of Persian writers for complexifying RA abstracts seems to be more lexicalized than that of international writers.

The findings of our study have several pedagogical implications. First, the present study has shown that expert non-native academic writers underused the most frequent type of noun phrase post-modifiers in academic writing, i.e. prepositional phrases, compared to international writers.

Since post-modifying prepositional phrases serve the dual purpose of raising explicitness in meaning and retaining the compressed writing style, special pedagogical attention should be paid to these widely-used lexico-grammatical features in academic writing. Data-driven learning (DDL) has been reported to be effective in teaching lexico-grammatical features, especially to graduate student writers who have already mastered the principles of text types such as expository or descriptive writing. One important thing that they need to acquire is the linguistic conventions of RAs. By providing authentic native-language RA corpora and concordancing computer programs as reference tools, DDL helps student writers promote language sensitivity, noticing, induction, and exemplar-based learning through such activities as hands-on projects of probability check of prepositional phrases, prepositional pattern of lemmas (basic word forms), concordancing high-frequency discipline-specific prepositional phrases, and so forth.

Second, lexical bundles constituted a considerable proportion of noun phrase modifiers in both corpora in our study. This highlights the importance of incorporating formulaic sequences into language curriculum, or EAP courses, because appropriate use of lexical bundles in academic registers, as our study showed, contributes to complex abstract writing. Noun phrase complexity through lexical bundles could "signal the text register" in "academic written genres" (Hyland, 2008a, p.5), possibly facilitating the predictability of the text to the readers.

Third, the present study indicated that international and non-native academic writers relied on different groups of lexical bundles for constructing noun phrase modifiers. Academic writing courses may benefit from compiling the list of target lexical bundles and the list of L2 writers' lexical bundles for developing curricula. Consciousness-raising tasks which underscore the differences in use of lexical bundle between native and non-native academic writers and contextually pinpoint any instances of overuse and underuse (Salazar, 2014) could be productive.

Like other studies, the current study had some limitations that need to be addressed. First, due to labor-intensive, and time-consuming process of qualitative check, which was part of our analysis, we employed a relatively small corpus of academic RA abstracts. It might be an interesting area of research for future studies to investigate less frequent features (like relative clauses with zero relativizers) which do not potentially lend themselves readily to systematic analysis in small corpora.

References

Ansarifar, A., Shahriari, H., & Pishghadam, R. (2017). 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*(1), 58-71.

Anthony, L. (2019). AntConc (Version 3.5.8) [Computer Software]. Waseda University. https://www.laurenceanthony.net/software

Appel, R., & Wood, D. (2016). Recurrent word combinations in EAP test-taker writing: Differences between high-and low-proficiency levels. *Language Assessment Quarterly*, *13*(1), 55-71.

Biber, D. (2003). Compressed noun phrase structures in newspaper discourse: The competing demands of popularization vs. economy. In J. Aitchison, & D. Lewis (Eds.), *New media discourse* (pp. 169–181). London: Routledge.

Biber, D. (2014). Using multi-dimensional analysis to explore cross-linguistic universals of register variation. *Languages in Contrast*, *14*(1), 7-34.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, *26*(3), 263-286.

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. *Language and Computers*, *26*, 181-190.

Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. *TESOL Quarterly*, *36*(1), 9-48.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371-405.

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, *9*(1), 2-20.

Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge: Cambridge University Press.

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, *45*(1), 5-35.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of spoken and written English*. London: Longman.

Casanave, C. P. (1994). Language development in students' journals. *Journal of Second Language Writing*, *3*(3), 179-201.

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology, 14*(2), 30-49.

Cheng, A. (2019). Examining the "applied aspirations" in the ESP genre analysis of published journal articles. *Journal of English for Academic Purposes, 38*, 36-47.

Cohen, J.W. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397–423.

Crawford, W. J., & Csomay, E. (2016). *Doing corpus linguistics*. New York, NY: Taylor and Francis Inc.

Halliday, M. A. K., & Martin, J. R. (1993). General orientation. In M. A. K. Halliday, & J. R. Martin (Eds.), *Writing science* (pp. 2-21). London: The Falmer Press.

Hamilton, R. (1994). Is implicational generalization unidirectional and maximal? Evidence from relativization instruction in a second language. *Language Learning, 44*(1), 123-157.

Harzing, A. W. K., & Van der Wal, R. (2008). Google scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics, 8*(1), 61-73.

Huckin, T. (2006). Abstracting from abstracts. In M. Hewings (Ed.), A*cademic writing in context: Implications and applications* (pp. 93-103). Birmingham: Birmingham University Press.

Hughes, R. (2005). *English in speech and writing: Investigating language and literature*. New York, NY: Routledge.

Hutter, J. A. (2015). *A corpus-based analysis of noun modification in empirical research articles in applied linguistics*. Dissertations and Theses. PDXScholar: Portland State University. http://pdxscholar.library.pdx.edu/open_access_etds.

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. London: Longman.

Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18*(1), 41-62.

Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*(1), 4-21.

Hyland, K., & Tse, P. (2005). Hooking the reader: A corpus study of evaluative that in abstracts. *English for Specific Purposes*, *24*(2), 123-139.

Jiang, J., Bi, P., & Liu, H. (2019). Syntactic complexity development in the writings of EFL learners: Insights from a dependency syntactically-annotated corpus. *Journal of Second Language Writing, 46*, 1-13.

Jiang, F. K., & Hyland, K. (2017). Metadiscursive nouns: Interaction and cohesion in abstract moves. *English for Specific Purposes*, *46*, 1-14.

Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal*, *102*(2), 333-349.

Lan, G., & Sun, Y. (2019). A corpus-based investigation of noun phrase complexity in the L2 writings of a first-year composition course. *Journal of English for Academic Purposes, 38*, 14-24.

Liu, X. L., & Liu, X. X. (2009). A corpus-based study on the structural types and pragmatic functions of lexical chunks in college English writing. *Foreign Languages in China, 6*, 48-53.

Lorés, R. (2004). On RA abstracts: From rhetorical structure to thematic organization. *English for Specific Purposes*, *23*(3), 280-302.

Loudermilk, B. C. (2007). Occluded academic genres: An analysis of the MBA thought essay. *Journal of English for Academic Purposes*, *6*(3), 190-205.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45*(1), 36-62.

Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, *29*(1), 16-27.

Manning, C. D. (2015). Computational linguistics and deep learning. *Computational Linguistics*, *41*(4), 701-707.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of annual meeting of the association for computational linguistics: system demonstrations* (pp. 55-60).

Marefat, H. & Rahmany, R. (2009). Acquisition of English relative clauses by Persian EFL learners. *Journal of Language and Linguistic Studies*, *5* (2), 21-48.

Myles, F. (2012). Complexity, accuracy and fluency the role played by formulaic sequences in early. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp.71–93). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Nesi, H., & Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education*. Cambridge: Cambridge University Press.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*(4), 555-578.

Öchsner, A. (2013). *Introduction to scientific publishing backgrounds, concepts, strategies*. Dordrecht: Springer.

Pallant, J. (2013). *SPSS survival manual*. McGraw-Hill Education: UK.

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*(1), 48-59.

Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory. In Richards, J. C. & Schmidt, R.W. (Eds.), *Language and Communication* (pp.191-227). London: Longman.

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*(2), 231-250.

Ravid, D., & Berman, R. A. (2010). Developing noun phrase complexity at school age: A text-embedded cross-linguistic analysis. *First Language*, *30*(1), 3-26.

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*(1), 37-47.

Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching* (Vol. 65). Amsterdam: John Benjamins Publishing Company.

Samraj, B. (2005). An exploration of a genre set: Research article abstracts and introductions in two disciplines. *English for Specific Purposes*, *24*(2), 141-156.

Song, M., & Chambers, T. (2014). Text mining. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact: Methods and practice* (pp. 215-234). Berlin: Springer.

Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication*, *33*(2), 149-183.

Stotesbury, H. (2003). Evaluation in research article abstracts in the narrative and hard sciences. *Journal of English for Academic Purposes*, *2*(4), 327-341.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Swales, J. (1996). Occluded genres in the academy: The case of the submission letter. In E. Ventola & A. Mauranen (Eds.), *Academic writing: Intercultural and textual issues* (pp. 45-58). Amsterdam: John Benjamins Publishing Company.

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*(2), 420-430.

Taylor, H., & Goodall, J. (2019). A preliminary investigation into the rhetorical function of 'I'in different genres of successful business student academic writing. *Journal of English for Academic Purposes*, *38*, 135-145.

Wei, N.X. (2007). Phraseological characteristics of Chinese learners' spoken English: Evidence of lexical chunks from COLSEC. *Modern Foreign Languages*, *30*, 280-291.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language & Communication*, *20*(1), 1-28.

Wu, X., Mauranen, A., & Lei, L. (2020). Syntactic complexity in English as a lingua franca academic writing. *Journal of English for Academic Purposes*, *43*, 1-13.

Appendix A: Hamilton's (1994) Classification of Relative Clauses Based on Processing Difficulty

Sentence type	Example
OS	Jerry likes the teacher who explained the answer to the class.
OO	A man bought the clock that the woman wanted.
SS	The man who needed a job helped the woman.
SO	The dog that the woman owns bit the cat.

Appendix B: Hierarchy of Difficulties in Relative Clauses as Represented in Hamilton's (1994) SO Hierarchy Hypothesis (SOHH)

Type	Example	Number of discontinuities
OS	She visited the professor$_i$ [$_S$ t$_i$ who opened the door].	1 (by relativization)
OO	She found the book that$_i$ [$_S$ the professor [$_{VP}$ wanted t$_i$]].	2 (1 by relativization and 1 within RC)
SS	The professor [who$_i$ [$_S$ t$_i$ opened the door]] wanted the book.	2 (1 by relativization, 1 by center embedding)
SO	The student [that$_i$ [$_S$ the professor [$_{VP}$ saw t$_i$]]] opened the door.	3 (1 by relativization, 1 by center embedding, and 1 within RC)

Note: S = sentential node. VP = verb phrase. t = *wh*-trace. i = co-index.