



The Journal of Asia TEFL

<http://journal.asiatefl.org/>

e-ISSN 2466-1511 © 2004 AsiaTEFL.org. All rights reserved.



Noun Phrase Complexity in EFL Academic Writing: A Corpus-Based Study of Postgraduate Academic Writing

Liming Liu

The Hong Kong Polytechnic University, China

Lan Li

The Hong Kong Polytechnic University, China

Noun phrase (NP) centered structures are distinctive syntactic devices in academic discourse. The commonly employed subordination-based complexity measures cannot adequately capture the development of syntactic complexity of noun phrases expected of advanced student academic writing (Biber, Gray, & Poonpon, 2011). Following the call for more research in this area (e.g. Lu, 2011, p. 57), the current study compared noun phrase complexity in two corpora: one is a corpus of MA dissertations written by Chinese EFL students and the other comprises published research articles in applied linguistics journals. The study examined overall noun phrase complexity using an automatic syntactic complexity analyzer and specifically identified features of one aspect of NP complexity: NP postmodification. The quantitative results were further contextualized in a textual analysis of excerpts from the two corpora for demonstrating the significance of NP complexity to the establishment of discourse coherence. Results of the analyses showed significant underdevelopment of NP postmodification complexity in student writing relative to published texts, meanwhile explicating the circumstances under which the difference is meaningful. Implications of the findings for the teaching of EFL academic writing were also discussed.

Keywords: syntactic complexity, noun phrase complexity, postmodification, EFL academic writing, contrastive corpus analysis

Introduction

The last two decades have seen a growing interest in research into advanced non-English-speaking writers as they are apprenticed and enculturated into the international academic community with English as the global *lingua franca* for international publication or pursuing research higher degrees (Uzunur, 2008). Research on advanced L2 (second language) writing includes examination of its linguistic and generic features, efforts made by novice writers for international publication, the dynamic change of author identity, and issues related

to cross-cultural understandings of plagiarism. Academic literacy develops on two parallel but related levels: linguistic literacy (Ravid & Tolchinsky, 2002) and discursive literacy (Blum-Kulka, 2004). The development of linguistic literacy concerns knowledge about the lexical and syntactic resources available in a language, whereas that of discursive literacy means a clear understanding of the linguistic and organizational variation across different genres. Therefore, acquiring these two types of literacy involves skilful manipulation of the lexical and grammatical resources in ways that conform to the conventions of the general academic discourse community and specific academic disciplines and text types (e.g. book reviews, abstracts, research proposals, and literature reviews).

For postgraduate programs with a research focus, academic achievement is most notably measured by students' ability to write up a qualified dissertation. To date, the majority of research on theses and dissertations has focused on such issues as organizational structures (e.g. Bunton 2002, 2005), use of citation (Sun, 2008), and writer identity development (e.g. Cadman, 1997), with little investigation into their lexicogrammatical features and the linguistic development of student writers. However, as a considerable number of research students whose L1 (first language) is other than English have to write dissertations in this academic *lingua franca*, linguistic challenges stand out as equally important. A common finding of research into the writing experiences, difficulties, and challenges as perceived by student authors indicates that students see their problems and obstacles more at the local lexicogrammatical level, i.e. a lack of skilful manipulation of proper vocabulary and syntax for the expression of ideas and arguments, than at the global, generic-structural level (Bitchener & Basturkmen, 2006; Lee & Tajino, 2008; Qian & Krugly-Smolka, 2008; Yeh, 2010). The present study aims to add to the existing literature of investigations into the grammatical features of advanced EFL academic writing at the postgraduate level.

Measuring Syntactic Complexity

An important indication of linguistic maturity in academic literacy development can be obtained by measuring the syntactic complexity (grammatical complexity) of written texts. Scholars in both first and second language writing research have employed syntactic complexity measures to gauge overall writing development (e.g. Beers & Nagy, 2011; Hunt, 1970; Ortega, 2003). Syntactic complexity has also been considered one of three useful indices of second language development, together with accuracy and fluency (Norris & Ortega, 2009).

Syntactic complexity in L1 and L2 writing research refers to the variety and sophistication of syntactic structures accessible to students in their writing tasks (Wolfe-Quintero, Inagaki, & Kim, 1998). A range of measures have been developed for examining L2 writers' syntactic complexity. Norris and Ortega (2009) identified five types of syntactic complexity measures (for both speaking and writing) after reviewing a series of research from 1965 to 2007. Two types of metrics were found to be frequently used for measuring L2 writing complexity: 1) length-based measures, e.g. mean length of sentence (MLS), mean length of T-unit¹ (MLT) and mean length of clause (MLC); 2) subordination-based measures, e.g. clauses per T-unit (C/TU). The prevalence of these two types of measures can be explained by the common belief that longer production units mean greater fluency and that more subordination leads to convoluted argumentation and reasoning. However, length-based measures and subordination-based measures may sometimes return conflicting quantitative claims on syntactic complexity. While the lengthening of a sentence can be realized by incorporating more dependent clauses, this is not the only way to extend a sentence. For example, in academic

¹ The T-unit is defined as "the shortest unit (the Terminable Unit, Minimal Terminable Unit, or T-Unit) which a sentence can be reduced to, and consisting of one independent clause together with whatever DEPENDENT CLAUSES are attached to it" (Richards & Schmidt, 2002, p. 566).

and professional writing it is not unusual to meet an extended sentence consisting of just one independent finite clause, like the following:

This modern and interactive ICOSA website can be accessed by Hong Kong tertiary learners interested in English independent learning. (Taken from ICOSA, Hong Kong Baptist University, <http://icosa.hkbu.edu.hk/grammar/nouns/complex-noun-phrases8/index.htm>)

This 19-word-long sentence does not show any subordination-induced complexity, and it would be extremely simple if gauged by subordination-based measures. In fact, the complexity of this sentence exists at the clause-internal phrase level: it is the use of noun phrases that gives rise to the lengthening of the sentence—*this modern and interactive ICOSA website* and *Hong Kong tertiary learners interested in English independent learning*.

Therefore, studies using measures that tap different aspects of complexity may produce varied results. Interpreting these results can be difficult without an explicit knowledge of different complexity types. In order to better illustrate syntactic complexity measurement, Ortega (2003) and Norris and Ortega (2009) have proposed a multidimensional construct interpretation of syntactic complexity, covering the measurement of length, subordination, and phrasal structures.

The strong emphasis on measuring subordination identified by Norris and Ortega reflects an inadequate understanding of the nature of syntactic complexity and its relationship with writing development. Researchers seem to prefer to think of complexity in terms of the amount of subordination contained in a potentially multi-clausal sentence or T-unit, overlooking other dimensions to the construct. The more important question is the compatibility of complexity measures with the text characteristics under investigation. Therefore, for the examination of academic writing, one needs to ask to what extent subordination alone represents the typical discourse characteristics of academic texts. Answers to this question can be found from both theoretical and empirical investigations of the linguistic features of academic writing.

Syntactic Complexity of Academic Writing

Theoretical discussions of the grammatical features of academic writing can be found in works by scholars from systemic functional linguistics (SFL). The SFL approach to academic language regards nominalization as a pivotal linguistic device that distinguishes formal technical writing from everyday spoken interaction and as a major source of syntactic complexity in writing (Halliday & Martin, 1993). Nominalization, i.e. verbs and adjectives converted to nouns through derivation, is a powerful device for packing and integrating information: what could be expressed in clauses is rendered into nominals, which can be further connected with or embedded into another clause, as the following example shows²:

1. *Since he drove the bus over-rapidly downhill, the brakes failed.* (10 words)
2. *His over-rapid downhill driving of the bus caused brake failure.* (10 words)

Thus, the two clauses in 1 are re-coded as noun phrases through nominalization and respectively become the subject and object in 2, connected by the causative verb *caused* according to the logical relationship between the two noun phrases. Although both the two sentences are 10 words in length, 2 can be argued to carry a higher degree of complexity than 1 because of the greater level of abstraction in the nominalized

² Examples are adapted from <http://folk.uio.no/hasselg/systemic/metaphor.htm>

propositions and the equally abstract causative relationship between the two propositions. Organizing meaning in this more “incongruent”, “metaphorical” way places a much greater cognitive challenge on the writer than does the more “congruent” clausal presentation (Halliday & Matthiessen, 1999, p. 227 ff.). This “highly information-packed, lexically dense” (Halliday, 1989, p. 87) style of writing is a major source of syntactic complexity of academic texts.

Corpus linguistics research has informed empirical investigation into the grammatical features of academic writing. Particularly revealing about the syntactic complexity of academic discourse is corpus-based analysis of linguistic variation between academic writing and other registers, notably informal conversation (e.g. Biber, 2006; Biber, Gray, & Poonpon, 2011). By examining the distribution of 28 grammatical features in academic research articles and conversation, Biber, Gray, and Poonpon reveal that syntactic structures significantly preferred in academic texts are “nonclausal features embedded in noun phrases” (p. 29), including postmodifying prepositional phrases in NPs (noun phrases), premodifying nouns in NPs, attributive adjectives in NPs, nonfinite relative clauses, and *wh*-relatives. By contrast, grammatical devices significantly more frequent in conversation are mostly finite clausal structures: finite adverbial clauses, verb + *that*-complement clauses, and verb + *wh*-complement clauses. In the 5-stage development index hypothesized by Biber et al. (2011), extensive phrasal embedding in the noun phrase with multiple prepositional phrases as postmodifiers has been proposed as the most sophisticated feature in stage 5. These findings empirically corroborate SFL’s theoretical claim that a highly nominalized and integrated grammar is the defining feature of academic writing.

In summary, both theoretical claims and empirical evidence have pointed to NP-centred structures as distinctive syntactic devices in academic discourse. Therefore, the commonly employed subordination-based complexity measures alone cannot adequately capture the development of syntactic complexity expected of advanced student academic writing, such as the postgraduate thesis. But this inadequacy has been increasingly heeded by L2 writing researchers, calling for additional measures more capable of gauging noun phrase based complexity features (e.g. Biber et al., 2011; Norris & Ortega, 2009).

However, a word of caution is needed for interpreting the extent of “complexity” of NP-based structures. In other words, one needs to ask the question how complex noun phrases should be in academic writing, given that there is in theory no limit to noun phrase complexity. Unrestrained and awkward use of pre- and postmodification for the mere sake of achieving noun phrase complexity would no doubt jeopardize smooth and effective comprehension of the text, and this should be clearly pointed out to students moving towards increased noun phrase complexity. On the other hand, emphasis on the prominence of NP-based structures as an information-packing device in academic writing should not make other equally important syntactic structures pale into insignificance. For example, although significantly less frequent in academic writing than in informal conversation (Biber et al., 2011, p.25), finite adverbial clauses are still important resources for explicating the logical and semantic relationships among clauses expressing already-complicated information and ideas. More importantly, adverbial conjunctions encode radically different discourse semantics in speech and academic writing (Schleppegrell, 1996), with those in spontaneous speech more readily serving as interactional and pragmatic markers and those in academic writing as markers of meaning relationships in text. Comparing the use of adverbial clauses only in frequency terms may conceal the more important stylistic differences. Apart from finite adverbial clauses, a range of other clause-integrating devices also figure prominently in academic writing, mostly participle clauses and verbless clauses reduced from their finite prototypes (Cho & Park, 2015; Granger, 1997). The process of phrase combining to achieve noun phrase complexity and that of clause combining to enhance logical-semantic coherence should co-operate for better text quality.

Research into Noun Phrase Complexity

Following calls for additional measures for syntactic complexity of academic writing, a growing body of research has emerged in the past few years that aims to chart the cross-sectional variability and longitudinal development of multidimensional complexity measures (e.g. Byrnes, 2009; Byrnes, Maxim, & Norris, 2010; Lu, 2011). It has been found that there exists a dynamic relationship between coordination and subordination-based complexity and phrasal complexity, exhibiting a trade-off between the two towards more advanced literacy levels (Byrnes et al., 2010; see also Ortega, 2003). In other words, as students' literacy grows, their writing is supposed to show decreased subordination and increased noun phrase related structures. Such a trade-off may very probably be the result of employing nominalization, by which process clauses are packed into noun phrases, thus reducing the amount of simple, oral-style subordination on the whole. Measures that gauge noun phrase complexity have also been found to significantly differentiate among writing samples of different proficiency levels; however, no measures for subordination density have been found to have this function (Lu, 2011). These findings further point to noun phrase complexity as a strong indicator of advanced literacy, deserving more attention from L2 writing researchers and practitioners alike.

This study aims to examine the clause-internal, noun phrase complexity of EFL academic writing at the postgraduate MA level by taking a corpus-based contrastive approach, with published writing as the frame of comparison. Apart from providing a general picture of noun phrase complexity of student writing, the study focuses on students' use of complex noun phrases with multiple postmodification—a device for highly elaborated complexification of the noun phrase (Biber et al., 2011)—and its distinctive role in constructing overall textual features.

The three research questions addressed by this study are:

1. To what extent does postgraduate academic writing differ from published research articles in overall noun phrase complexity?
2. To what extent does postgraduate academic writing differ from published research articles in the complexity of noun phrase postmodification?
3. In what ways do differences in noun phrase complexity (if any) affect the textual features of academic writing?

Methodology

The study addressed the three questions taking a corpus-based approach. The first question was answered by using an automatic computational tool to calculate and compare three measures of noun phrase complexity between the two groups of writers, while the second was answered by closely looking at the use of postmodification in a sample of complex noun phrases (CNPs)³. Detailed textual analysis of student writing and published writing was performed to answer the third question.

Data

Two corpora have been compiled for contrastive analysis: the MA dissertation corpus (hereinafter referred to as the MDC) and the published research article corpus (hereinafter referred to as the PRC). The MDC consists of 70 dissertations written by Chinese EFL postgraduate students studying applied linguistics in

³ In this study a complex noun phrase is defined as a noun phrase with at least one postmodifier.

China, with a total size of 1,019,243 running words (tokens). The samples were collected from the China Masters' Theses Full-text Database (CMFD), which "collects dissertations with distinction from more than 693 institutions offering master degree programs in China since 1984" (CNKI, 2014, authors' own translation). The PRC comprises 129 empirical research articles taken from six top journals in applied linguistics, totalling 971,726 words. The PRC features authorship covering a variety of L1 backgrounds, which means that this study is not a comparison between non-native and native speakers of English but rather focuses on the variation between advanced EFL students and professional writers with respect to noun phrase complexity.

The L2 Syntactic Complexity Analyzer

The study used the L2 Syntactic Complexity Analyzer (L2SCA) developed by Lu (2010) to automatically compute three measures of overall noun phrase complexity for each text in the two corpora. The three measures include a length-based metric of noun phrase complexity—mean length of clause (MLC), and two specific noun phrase complexity metrics—number of complex nominals per clause (CN/C) and number of complex nominals per T-unit (CN/T).

The L2SCA targets advanced second language writing although it "can be used on any English text in which sentence completeness is not a major problem" (Lu, personal communication, 26 June, 2013). Five types of syntactic complexity measures (altogether 14 measures) are included in the analyzer: length of production unit, sentence complexity, subordination, coordination, and particular structures. The system produces meaningful results only when grammatical completeness of the target texts is not a major problem, i.e. largely free of grammatical errors. It might be argued that grammatical errors are a major feature of second language texts and would therefore distort syntactic parsing outcomes, thus rendering complexity values falsified. However, the major linguistic problem with advanced second language writing, e.g. MA dissertations, resides not so much in grammaticality as in appropriate manipulation of idiomaticity and stylistics in ways that are preferred in the academic discourse (Lu, 2010).

The L2SCA first calls the Stanford parser (Klein & Manning, 2003) to segment an input text into individual sentences, which are then tokenized and part-of-speech (POS) tagged using the same parser; the input should be plain text without any form of annotation. The parsing process generates sequences of parse trees, which the L2SCA asks Tregex (Levy & Andrew, 2006) to query in order to retrieve and count the production units and syntactic structures according to Tregex patterns manually defined for the units and structures. Finally, the L2SCA calculates the fourteen syntactic complexity indices based on the counts by Tregex. Lu (2010) reported a degree of system-annotator reliability for identifying the production units and syntactic structures involved in the fourteen measures, with F-scores ranging from .830 for complex nominals to 1.000 for sentences. The complexity scores computed by the annotators and the system also achieved strong correlations, ranging from .834 for CP/C to 1.000 for MLS. The analyser can process a single text or a collection of texts saved in a folder, i.e. a corpus composed of individual texts. The results of the 14 syntactic complexity measures generated by the L2SCA can be easily sent to Excel or SPSS for statistical analyses. A detailed, step-by-step procedure for how to use the analyser is illustrated in Lu (2014).

The definition of "clause" adopted in the L2SCA follows the tradition of writing research where the term only refers to finite clauses, with non-finite elements (e.g. infinitives and participial structures) regarded as verb phrases. According to Lu (2010, p. 483), "complex nominals comprise (i) nouns plus adjective, possessive, prepositional phrase, relative clause, participle, or appositive, (ii) nominal clauses, and (iii) gerunds and infinitives in subject position". It can be seen that only the first component has relevance for complex noun phrases with extended postmodification (the most elaborated stage of syntactic complexity hypothesized by Biber et al., 2011). Therefore, only the two complex nominal measures have been selected to

represent overall noun phrase complexity and detailed manual analysis was performed to investigate postmodification complexity.

Analytical Procedure

After the calculation of the three measures for each text in the two corpora, an independent samples *t*-test was run to see if there was a significant difference in the three measures between the two corpora using SPSS 20.0, with the significance level set at $p < .01$.

In this study, multiple postmodification was not just confined to the consecutive use of prepositional phrases as postmodifiers, but also included other types of grammatical structures, as in the following:

an understanding of language as an ideational signifying system that plays a central role in how we understand ourselves and the world.

In this example postmodifiers include two prepositional phrases and a relative clause. Since the three measures for noun phrase complexity included in the automatic L2SCA does not directly represent the extent of multiple postmodification,⁴ this feature can only be examined by hand. Therefore, in order to limit the data to a manageable size, the present study only analyzed complex noun phrases whose first postmodifier was prepositional phrases headed by *of*. This is because multiple postmodification is most possible with the first postmodifier being a prepositional phrase and *of*-phrases are the most frequent prepositional phrase in English (Biber et al., 1999, pp. 635-642). Finally, complex noun phrases expanded from three head nouns were examined—*understanding*, *analysis*, and *lack*, in that they have comparable frequencies across the two corpora when postmodified by *of*-phrases. Although fairly limited in scope, the sample of data thus selected still worked well for a close scrutiny and comparison of postmodification across the two corpora.

The structural types of postmodifiers identified in this study are based on those provided by Biber et al. (1999, pp. 604-605), including 1) finite relative clauses, 2) non-finite postmodifying clauses, 3) noun complement clauses, 4) prepositional phrases, 5) adverb (phrases), 6) adjective (phrases), and 7) reflexive pronouns. The study only identified restrictive postmodifiers, disregarding non-restrictive postmodification.

The specific aspects of postmodification examined in this study were: 1) length of overall postmodification; and 2) depth of multiple postmodification. Length of the overall postmodification is gauged by counting the number of words from the first word of the first postmodifier (i.e. the preposition *of*) until the last word of the last postmodifier. Depth of postmodification is determined by counting the number of postmodifiers expanding from the head noun, with each postmodifier understood as constituting one “level” of postmodification. Roughly following Biber et al.’s notation (1999, p. 576), multiple levels of postmodification is exemplified as below, with the postmodifiers marked by [] and levels of postmodification by the subscript numerals _{1, 2, 3} etc.:

a logical analysis [₁of content representation [₂of test items [₃demonstrated in the table [₄of specifications₄]₃]₂]₁].

Greater postmodification depth is indicated by more postmodification levels. However, if a finite clause (and also some non-finite *wh*-clauses) appears in the postmodification of a CNP, e.g. as the object of a preposition phrase or as a complement clause, any noun phrases with postmodification within that clause would not be further analyzed as constituting an extra depth level. For example, in the following CNP,

⁴ For example, complex noun phrases with multiple postmodifiers would be parsed as one instance of complex nominal by the L2SCA.

our understanding of how the investigation of questions, including their grammar, as interactional products can be a catalyst to comprehending the connection between grammar and social organization

since the two complex noun phrases *the investigation of questions, including their grammar, as interactional products* and *a catalyst to comprehending the connection between grammar and social organization* occur within a finite *how*-clause, they are not seen as expanding from the head noun *understanding*, hence their exclusion from depth analysis.

Extraction of the complex noun phrases was facilitated by WordSmith Tools 5.0 (Scott, 2008) but the boundary of the complete postmodification and that of each level of postmodifier for each instance of CNP were individually determined by the researchers. Subsequently, the length of the complete postmodification was counted for each CNP and saved in SPSS 20.0 for comparison of mean lengths; and the number of levels of postmodifier for each CNP was figured out and saved in Microsoft Excel for comparing depth of postmodification.

Results and Discussion

Overall Noun Phrase Complexity

The L2SCA was used to compute scores of the three noun phrase complexity measures, i.e. mean length of clause (MLC), complex nominals per T-unit (CN/T), and complex nominals per clause (CN/C). Table 1 summarizes the descriptive mean scores of the three measures across the two corpora. As can be seen, the means of the PRC outnumber those of the MDC on all three measures, which means that student writers produced shorter clauses and fewer complex nominals per clause and T-unit than the published writers. It can also be seen that the PRC has a greater standard deviation than the MDC for each of the three measures, indicating that a stronger variability within the research article corpus than the dissertation corpus. The smaller standard deviation in the MDC suggests that scores of each measure tended to cluster more closely to the mean score compared with the scores of the PRC. In other words, the mean scores of the three measures for the MDC are more truly representative of each text in the corpus due to the lower standard deviation. Thus, the lower mean scores of the three measures accompanied with smaller standard deviations in the MDC indicate that the 70 texts in the corpus are more or less equally lower on the three measures in comparison with the PRC.

Information on statistical significance of the differences is presented in Table 2. With significance level set at $p < .01$, statistics from the *t*-test show that a significant difference existed between the two corpora for all of the three measures. In order to examine the effect size, or the strength, of the significant differences, the commonly employed effect size coefficient Cohen's *d* was computed for the significant difference of each measure.⁵ Values of Cohen's *d* ranged from -0.562 to -0.772, indicating a moderate to high practical significance (Cohen, 1988). The results suggest a much lower degree of overall noun phrase complexity in MA dissertations than in published research articles, and testify to the ability of noun phrase complexity as a strong indicator of syntactic complexity development in academic writing given that student writers and published writers should be considered as representing two distinct developmental stages (Biber et al., 2011).

⁵ The calculation of Cohen's *d* was performed using an online effect size calculator offered by UCCS at <http://www.uccs.edu/~lbecker/>.

TABLE 1

Means of Three Noun Phrase Complexity Measures between the Two Corpora

Measures	Corpus	N	Mean	Std. Deviation
MLC	MDC	70	12.140	1.370
	PRC	127	13.192	1.717
CN/T	MDC	70	2.951	.475
	PRC	127	3.410	.619
CN/C	MDC	70	1.698	.223
	PRC	127	1.863	.313

TABLE 2

Independent Samples t-Test of Noun Phrase Complexity Measures

Measures	t	df	Sig. (2-tailed)	Mean Difference	Cohen's d
MLC	-4.410	195	.000	-1.052	-0.632
CN/T	-5.393	195	.000	-0.459	-0.772
CN/C	-3.921	195	.000	-0.166	-0.562

Note. Significance level is set at $p < .01$.

Returning to the noun phrase complexity differences in the present study, it is important to note the magnitudes at which between-corpus differences reached statistical significance for the three measures, i.e. 1.052 for MLC, 0.459 for CN/T, and 0.166 for CN/C. Take the magnitude for MLC for example, a mean difference of approximately one word per clause between the two corpora would contribute to statistical significance. Therefore, the challenge in qualitatively interpreting the difference is obvious in that it would be difficult to imagine what a mean difference of one word per clause signifies in the actual writing process and how it could shed meaningful light on teaching academic writing. Furthermore, as previously noted, the L2SCA regards a complex noun phrase with multiple layers of postmodification as a single case of “complex nominal”. This has important implications for examining noun phrase complexity. A text composed of more extended noun phrases (i.e. those with multiple postmodifiers) may end up having a similar, or even smaller, number of “complex nominals” than one composed of noun phrase with much fewer postmodifiers, based on the calculation of the L2SCA. However, the two texts deserve different interpretations of noun phrase complexity, with the one featuring more extended nominals being more cognitively demanding, and hence, more complex (Cummins, 1984). For this reason, a more insightful comparison of noun phrase complexity has to consider the extent to which noun phrases condense information, most notably through multiple postmodifiers.

Postmodification in the Noun Phrase

The examination of multiple postmodification was based on a sample of complex noun phrases (CNPs) with their first layer of postmodifier being an *of*-preposition phrase and expanded from three head nouns—*understanding*, *analysis*, and *lack*. The selection of these nouns as the head words from which complex noun phrases expand arises from the following two considerations: 1) they have relatively similar frequencies in the two corpora when postmodified by an *of*-preposition phrase (see Table 3); and 2) all three nouns are included in Gardner and Davies' (2013) *Academic Vocabulary List*.

TABLE 3
Occurrences of CNPs for Each Head Noun across Corpora

	MDC	PRC
<i>understanding</i>	232	251
<i>analysis</i>	272	275
<i>lack</i>	152	224

In addition, selecting more than one head noun allows for a wider representation of complex noun phrases since nouns could differ in determining the complexity of postmodification due to their inherent semantic properties. For example, the postmodification of *effect* and *relationship* would be potentially complex because these two nouns semantically entail a postmodifying complex, i.e. *effect of...on...* and *relationship between...and....* The reason for limiting the first postmodifier to an *of*-phrase is that in English preposition phrases are the most frequent postmodifier type and *of* is the most frequent preposition (Biber et al., 1999). Two criteria were used as measures of complexity of postmodification: 1) length of postmodification; and 2) depth of postmodification.

Length of Postmodification

First, the mean lengths of postmodification for CNPs with each head noun across the two corpora were calculated and a *t*-test was run to test statistical significance of the length difference between the two corpora. Statistics for the mean lengths and significance tests (together with the effect size Cohen's *d*) are listed in Table 4.

TABLE 4
Comparison of Mean Lengths of Postmodification

	MDC	PRC	Cohen's <i>d</i>
<i>understanding</i>	5.59**	7.10	-0.355
<i>analysis</i>	5.93*	7.04	-0.269
<i>lack</i>	5.59*	6.71	-0.269

*. $p < .01$
**. $p < .001$

As can be seen, there was a significant difference between the two corpora in the length of postmodification for CNPs with each head noun, with the magnitudes at which length differences reached significance ranging from 1.12 to 1.51 words. This result is consistent with the previous finding that the two corpora are significantly different in terms of overall noun phrase complexity, thus establishing the mean length of noun phrase postmodification as a reliable measure for noun phrase complexity. However, as the mean length scores only show the idealized central tendency for each set of complex noun phrases, it does not necessarily represent the actual scenario of postmodification complexity. For example, occupying different information positions in the flow of discourse, there were in both corpora instances of very short postmodification, such as

Iris's and Yumin's understanding of plagiarism (2 words, PRC)
the understanding of the pun (3 words, MDC),
and extremely lengthy cases—sometimes as long as nearly 20 words, such as
an initial understanding of the extent to which the students believed in the existence of FLA in foreign language learning (17 words, PRC)

a deeper understanding of metacognitive strategies as a kind of higher executive strategies than the other two kinds of strategies to plan (19 words, MDC).

Thus, to go beyond the comparison of postmodification length, it would be necessary to know the proportion of longer instances of postmodification as a complementary interpretation of the significant length differences of postmodification between the two corpora.

TABLE 5
Proportion (%) of Longer CNPs between the Two Corpora

	MDC	PRC	z-score
<i>understanding</i>	21.2**	34.8	-3.306
<i>analysis</i>	25.7*	36.7	-2.773
<i>lack</i>	20.3*	34.8	-3.026

*. $p < .01$

** . $p < .001$

For pure comparison's purpose, an arbitrary cut-off point of 8 words was set up to determine postmodification greater than 8 words as "longer", and their proportion between the two corpora was compared. Table 5 presents the comparison of the proportion of longer postmodification between the two corpora. The statistics indicate that there was a lower proportion of heavier postmodification in the MDC than in the PRC, with differences reaching statistical significance according to the z-test.⁶ While the proportion of longer postmodification in the MDC ranged from around 20% to 26%, that in the PRC from 35% to 37%, with the difference being significant for CNPs headed by all three nouns. Thus, not only did CNP postmodification in the PRC have a greater mean length than that in the MDC, but there was also a greater amount of longer postmodification in the PRC.

Depth of Multiple Postmodification

The depth of multiple postmodification is represented by the proportion of consecutive postmodifiers in the selected CNP samples. Table 6 presents proportional information on CNP postmodification depth. As is shown in Table 6, for CNPs headed by all three nouns both the two corpora have reached Level 4 and for CNPs headed by two nouns (i.e. *analysis* and *lack*) both the two corpora have reached Level 5. However, postmodification from Level 5 onwards was so rare a phenomenon for both corpora that a comparison between the two groups of writers in this regard may not offer much revealing insight. Therefore, a general picture of postmodification depth across the two corpora can be outlined from a close examination of the distribution of postmodifiers at Levels 2, 3, and 4.

⁶ The z-score test for two population proportions is used for examining whether two populations or groups differ significantly on some single (categorical) characteristic – longer postmodification in the case of this study. The z-scores were computed using a calculator offered by socscistatistics.com at <http://www.socscistatistics.com/tests/ztest/Default.aspx>.

TABLE 6
Proportion of Multiple Postmodifiers

		<i>understanding</i>		<i>analysis</i>		<i>lack</i>	
		PRC	MDC	PRC	MDC	PRC	MDC
Max. no. of levels		5	4	6	5	8	5
Instances of CNPs		251	232	275	272	224	152
Level 2	Instances	116	81	129	113	119	67
	Proportion	46.2%	34.9%	46.9%	41.5%	53.1%	44.1%
Level 3	Instances	31	16	42	38	55	25
	Proportion	12.4%	6.9%	15.3%	14.0%	24.6%	16.4%
Level 4	Instances	8	4	11	11	14	6
	Proportion	3.2%	1.7%	4%	4.0%	6.3%	3.9%
Level 5	Instances	2	--	1	2	6	2
	Proportion	0.8%	--	0.4%	0.7%	2.7%	1.3%

Note. The percentages have been rounded to the first decimal place.

It can be seen that for CNPs headed by each of the three nouns, the PRC saw a higher proportion of CNPs with multiple postmodifiers at all three levels (2, 3, and 4) as compared with the MDC, except for CNPs headed by *analysis* at Level 3 and 4, which exhibited an equal or almost equal proportion across the two corpora. For example, for CNPs headed by *understanding*, 46.2% of those in the PRC had a Level-2 postmodifier while only 34.9% in the MDC did; and 12.4% in the PRC had a Level-3 postmodifier compared to 6.9% in the MDC. It can also be observed that across the three levels of postmodifiers, Levels 2 and 3 exhibited a larger gap in their presence between the two corpora than did Level 4. Generally speaking, the depth of multiple postmodification in the PRC has shown to be greater than that in the MDC, with the major gap in postmodification depth attributed to postmodifiers at Level 2 and 3, and Level 4 only displaying a marginal difference.

The reason why postmodification depth for CNPs headed by *analysis* has exhibited a somewhat smaller differentiation between the two corpora could be that student writers attempted to use the phraseology *analysis of* to describe their research in great detail. Thus, students' intimate familiarity with the content of their own studies (e.g. object of analysis, method of analysis, aim of analysis, etc) has contributed to the production of the relatively expansive CNPs, such as the following rather full-blown CNP headed by *analysis* from the student corpus:

detailed text analysis [1of the authentic student pieces [2in the massive database [3built by the "Experiencing English Writing" research group3]2]1].

The investigation into the complexity of noun phrase postmodification has so far demonstrated that complex noun phrases in the student corpus feature a lower degree of postmodification length and depth in comparison with the expert corpus. This finding is consistent with the results for overall noun phrase complexity as measured by MLC, CN/C, and CN/T. With respect to depth of multiple postmodification, two observations need to be noted: 1) even in the expert corpus multiple postmodification mostly stopped at the second level, i.e. most (around 50%) of the CNPs in the PRC had only two consecutive postmodifiers; and 2) the student writers trailed the expert writers mainly at Levels 2 and 3, with postmodification from Level 4 onward being a relatively insignificant presence. These findings have important implications for EAP pedagogy, indicating that, despite a crucial feature of advanced literacy, multiple postmodification should not be over-emphasized as having extreme complexity and that instructional attention needs to be focused on raising students' awareness of elaborating information within three levels of postmodification, and especially within the first two.

Textual Evidence of Quantitative Findings

To answer the third research question, this section provides textual evidence of quantitative results of overall noun phrase complexity and postmodification complexity. The contrasts in clause length and noun phrase density as well as the kind of preciseness and efficiency in packing up information between student and published writing can be illustrated by the following extracts taken from the introduction part of an MA thesis and a research article:

- (a) ***The information expressed in either speaking or writing should be clear enough to be understood by certain structure. It is not only the requirement of everyday communication, but also the requirement of any other written text in books, newspapers, magazines and so on. This is because a special property, which is called ‘coherence’, is lying in language. Coherence is an important notion in discourse analysis, and it exists everywhere in our daily life. It is the semantic or functional relation that implies in the deep structure of discourse that produces coherence.*** [91 words] (Text 42, MDC)
- (b) ***Research has shown just how significant proficiency in the main language of the destination country is to the economic integration of immigrants in Canada and the United States. Specifically, an increase in proficiency in this main language enhances the effects of earnings from education and the labour experience of the country of origin. In the United States, immigrants who are proficient in English have earnings approximately 17% higher than immigrants with limited English proficiency. A report by Statistics Canada using data from the 2006 census highlights the economic benefits of speaking only English or French in the workplace.*** [98 words] (Text 6, PRC; citations in the text have been deleted)

Extract (a) consists of 91 words in 9 clauses and extract (b) includes 98 words in 6 clauses, thus resulting in the former being less complex than the latter in terms of mean length of clause (MDC)—10.1 vs. 16.3 words per clause. What contributes to the clauses in (b) packing in more information is the more condensed use of complex noun phrases with postmodification than (a)—8 CNPs in 6 clauses vs. just 4 in 9. That is, there is on average 1.3 CNPs per clause in (b), but only about 0.5 CNP per clause in (a). Even without calculation, such a distinction in the density of CNPs and the length of postmodification between the two extracts can easily be perceived from the visual impact of the highlighted text (CNPs are highlighted in bold with postmodification underlined). Moreover, (b) features more instances of multiple postmodification than (a) (5 vs. 2). It is also striking to note that 4 of the 8 CNPs in the PRC extract occupy the syntactic slot of grammatical subject while only 1 does so in the MDC extract, echoing the observation by Vande Kopple (1994, p. 534) that “grammatical subjects in scientific discourse are markedly long” and Master’s (1991) finding that inanimate-abstract subjects with active verbs prevalent in scientific writing presents great challenge for L2 student writers. By contrast, the grammatical subjects in the MDC extract are mostly (60%) deictics (i.e. *it*, *this*, and *which*), a linguistic feature closely related to context-dependent text types with considerable audience involvement, for example oral interaction (Biber, 1988). On another note, the perceived grammatical unsophistication of (a) also derives from the use of clausal coordination in two sentences (i.e. *it is not only... but also...*, and *Coherence is... and it...*) to achieve parallelism between clauses as a rhetorical device in text development. However, the frequent use of parallel constructions, argues Kaplan (1966, p. 8), “would stand in the way of clear communication”. In fact, parallelism is also implicitly constructed among the sentences in (a) in that the grammatical subjects of sentences 2-4 are just simple repetitions of an idea expressed in the preceding sentence. While such a practice may seemingly observe the information distribution principle from old to new (Quirk et al., 1985), simply repeating the old information without any further specifications related

to the context under investigation only contributes to constant logic break-off and engenders tedium for reading. In sharp contrast, the CNP-constructed subjects in (b) are highly effective in bringing the given information to bear upon its related elements as discerned by the author.

A more careful observation of the two extracts reveals that the use of apparently rather loose syntactic structures in (a) is due to student writers' failure to establish the logical relationships among ideas expressed in the clauses and sentences; the kind of parallelism in extract (a) would obscure the logical connection among ideas, the identification of which is only left to the devices of the reader. Establishing such relationships requires clauses encompassing a single idea to be reduced into constituent parts of a higher-level clause through nominalization and other integrating devices (see Tannen, 1982, p. 39 ff.). The crucial mechanism for bringing these reduced constituents into meaningful contact hinges upon the effective selection of lexical resources capable of fulfilling the predicate system in the clause—typically through lexical verbs and linking verbs plus predicate nouns, adjectives, and prepositions. The implementation of this syntactic-discourse strategy is clearly discernable in extract (b), most notably in the frequent use of lexical verbs to connect nominalized subjects and objects, including *Research has **shown** that..., an increase in proficiency...**enhances** the effects of..., and A report by...**highlights** the economic benefits of...* While syntactically very simple, this basic SVO structure with heavy nominal participants would nevertheless be highly demanding of the writer's cognitive ability to envisage the logical and semantic relationships among the nominal components and of his/her lexical knowledge to demonstrate those relationships. In this sense, the MDC extract virtually devoid of this discourse feature may suggest that the student writer was deliberately avoiding the cognitive challenge of integrating and synthesizing information or simply his/her failure to identify the abstract relations among the propositions. Either way, however, the need for explicit instruction on this syntactic-discourse strategy is obvious.

However, it should be recalled that there are also highly complex cases of noun phrases in the student corpus as exemplified above (the CNP headed by *analysis*), particularly when it comes to students describing in detail the methodological procedures of their own studies. Admittedly, in contrast with the extract given above, some MA dissertations in the corpus have already shown highly mature noun phrase complexity, not exhibiting a major difference from expert writing, such as the following:

According to the analysis of the pre-test and the post-test, the comprehensive language ability of the students in EG was improved after the experiment in contrast with that of CG which almost remained the same. Via analysis of the scores of the students in different groups who were of different levels (group A, B, C, D), it can be inferred that the approach of incorporating lexical chunks into English language teaching had great positive effect on all the students except the best students whose average score increased only a little in contrast with those of the rest students. (Text 23, MDC)

One issue is that the presence of noun phrase complexity is not stable across the writings by the MA students, with some students demonstrating a much greater complexity than others. On the other hand, one needs to bear in mind that an important aspect of syntactic maturity concerns the author's ability to vary sentence length by mixing longer and shorter sentences as recommended by many technical writing guides, with shorter sentences usually functioning to grab readers' attention, such as the topic sentence in the following extract from the PRC:

There are two key issues that raise further discussion. First, there were few consistent correlations between linguistic scores and the full range of WM scores. This issue reflects conflicting results from other L2 WM research as well as suggestions that WM may be task-specific...

Therefore, the important message for EAP writing instruction is that despite the fact that noun phrases with complex postmodification permeate in published academic writing, they definitely do not characterize every single sentence in the unfolding of the text, subject to the requirement of rhetorical function and stylistic diversity.

Conclusion and Pedagogical Implications

Complex noun phrases are one of the most distinctive clause-level features in English academic discourse (Biber et al., 2011; Lu, 2011; Norris & Ortega, 2009). This study compared writings by advanced Chinese EFL students and published writers and found a significantly lower noun phrase complexity in the student corpus than in the published corpus on all three NP complexity measures: mean length of clause; complex nominals per T-unit; and complex nominals per clause. Similarly, complexity in the postmodification of complex noun phrases was also weaker in the student corpus in terms of the length and depth of postmodification. However, more meaningful interpretation of the findings is needed, especially with regard to the logic and interpretive meaning of postmodification. Multiple postmodification in published research articles was more frequent, but not necessarily with extreme complexity, as shown in the findings of this study. It has been found that most complex noun phrases in the expert corpus only have two or three consecutive postmodifiers. This can be used as a practical guide for raising students' awareness of the degree of complexity of multiple postmodification common in published academic writing.

The textual analysis of a student writing extract has evidenced the quantitative finding that student writing featured shorter clauses, fewer complex nominals, and a weaker NP postmodification than expert writing. Meanwhile, it shows how an underdeveloped noun phrase complexity contributes to a motion of discourse that lacks coherence, logical connection among ideas, and authorial evaluation of the matters under investigation. The impact of noun phrase complexity on the construction of textual coherence as illustrated in the textual analysis immediately echoes the finding by McNamara, Crossley, and McCarthy (2009) that syntactic complexity as measured by the number of words before the main verb is one of the three most predictive indices of L1 writing quality (along with two measures of lexical complexity), because the elements appearing before the main verb are mostly NP subjects and NP-based adverbials (e.g. preposition phrases and non-finite adverbial clauses). Although the conclusion cannot be drawn that noun phrase complexity directly leads to better coherence and higher writing quality, it is safe to claim that skillful manipulation of NP-related constructions constitutes an important textual features of high quality writing.

The findings of the study suggest the importance of enhancing student writers' noun phrase complexity as a potentially rewarding area of language-focused instruction. While EFL research students may be able to produce a style that reaches the basic threshold of academic writing, explicit instruction on the grammatical features and their discourse functions of this genre could help students progressively develop a more elaborated style. Efforts can be made to raise students' awareness of the saliency of NP-based structures for realizing different clause elements (e.g. subject, object, complement, and adverbial) in academic writing. Students can also be encouraged to compare their own writing with expert writing, focusing on how they use NP-based structures differently from expert writers. As students' awareness of NP-based structures grows, they can proceed to revise their own writing with the guidance of teachers, focusing on possibilities where loosely-organized structures can be reshuffled into more information-intensive NP-based constructions. Again, teachers and students should be aware of the danger of magnifying postmodification complexity as it may hamper smooth comprehension of ideational meaning and make the flow of discourse rather awkward. When emphasizing the importance of noun phrase complexity, teachers need to demonstrate how longer and shorter constructions mingle in a piece of writing to augment stylistic variety.

The Authors

Liming Liu is a PhD candidate in the Department of English at The Hong Kong Polytechnic University. He is currently engaged in a corpus-based project that examines the lexical and grammatical complexity of advanced EFL students' academic writing. He has given paper presentations at international conferences and a recent paper has been published in the *Proceedings of ASIALEX 2015 Hong Kong*.

Department of English
The Hong Kong Polytechnic University
Kowloon, Hong Kong
Tel: +852 2766 7978
Email: liming-leo.liu@connect.polyu.hk

Dr. Lan Li is an Associate Professor in the Department of English at The Hong Kong Polytechnic University, and the Fellow of Chartered Institute of Linguists, UK. She holds MPhil and PhD degrees in Applied Linguistics from the University of Exeter, UK. She has been teaching English at the university level for over 20 years tutoring undergraduates, postgraduates, and professional trainees. Her research interests and publications cover lexicology, lexicography, professional communication, corpus linguistics and sociolinguistics.

Department of English
The Hong Kong Polytechnic University
Kowloon, Hong Kong
Tel: +852 2766 7978
Email: lan.li@polyu.edu.hk

References

- Beers, S. F., & Nagy, W. E. (2011). Writing development in four genres from grades three to seven: Syntactic complexity and genre differentiation. *Reading and Writing, 24*(2), 183-202.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge, UK: Cambridge University Press.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
- 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). *The Longman grammar of spoken and written English*. London: Longman.
- Bitchener, J., & Basturkmen, H. (2006). Perceptions of the difficulties of postgraduate L2 thesis students writing the discussion section. *Journal of English for Academic Purposes, 5*(1), 4-18.
- Blum-Kulka, S. (2004). The role of peer interaction in later pragmatic development. In R. A. Berman (Ed.), *Language development across childhood and adolescence* (pp. 191-210). Amsterdam: John Benjamins.
- Bunton, D. (2002). Generic moves in PhD thesis introductions. In J. Flowerdew (Ed.), *Academic discourse* (pp. 57-75). London: Pearson Education.
- Bunton, D. (2005). The structure of PhD conclusion chapters. *Journal of English for Academic Purposes, 4*(3),

- 207-224.
- Byrnes, H. (2009). Emergent L2 German writing ability in a curricular context: A longitudinal study of grammatical metaphor. *Linguistics and Education*, 20(1), 50-66.
- Byrnes, H., Maxim, H., & Norris, J. M. (2010). Realizing advanced foreign language writing development in collegiate education: Curricular design, pedagogy, assessment [Monograph]. *Modern Language Journal*, 94(S1).
- Cadman, K. (1997). Thesis writing for international students: A question of identity? *English for Specific Purposes*, 16(1), 3-14.
- Cho, D. W., & Park, J. (2015). Supplementive participle clauses in science journal papers by Korean graduate students: A corpus-based study for pedagogical purposes. *The Journal of Asia TEFL*, 12(4), 1-36.
- CNKI. (2014). China Master Theses Full-text Database. Retrieved February 02, 2016, from <http://epub.cnki.net/kns/brief/result.aspx?dbPrefix=CDMD>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cummins, J. (1984). Wanted: A theoretical framework for relating language proficiency to academic achievement among bilingual students. In C. Rivera (Ed), *Language proficiency and academic achievement* (pp. 2-19). Clevedon: Multilingual Matters.
- Gardner, D., & Davies, M. (2013). A New Academic Vocabulary List. *Applied Linguistics*, 35(3), 305-327.
- Granger, S. (1997). On identifying the syntactic and discourse features of participle clauses in academic English: native and non-native writers compared. In Aarts, J., de Mönnink, I., & Wekker, H (Eds.) *Studies in English language and teaching* (pp. 185-198). Amsterdam & Atlanta: Rodopi.
- Halliday, M.A.K. (1989). *Spoken and written language*. Oxford, UK: Oxford University Press.
- Halliday, M.A.K., & Martin, J.R. (Eds.). (1993). *Writing science: Literacy and discursive power*. Washington, D.C: The Falmer Press.
- Halliday, M.A.K., & Matthiessen, C.M.I.M. (1999). *Construing experience through meaning: A language-based approach to cognition*. London: Continuum.
- Hunt, K. W. (1970). Do sentences in the second language grow like those in the first? *TESOL Quarterly*, 4(a), 195-202.
- Kaplan, R. B. (1966). Cultural thought patterns in inter-cultural education. *Language Learning*, 16(1-2), 1-20.
- Klein, D., & Manning, C. D. (2003). Fast exact inference with a factored model for natural language parsing. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems 15* (pp. 3-10). Cambridge, MA: MIT Press.
- Lee, S. C. N., & Tajino, A. (2008). Understanding Students' Perceptions of Difficulty with Academic Writing for Teacher Development: A Case Study of the University of Tokyo Writing Program. *京都大学高等教育研究*, 14, 1-11.
- Levy, R., & Andrew, G. (2006). Tregex and Tsurgeon: Tools for querying and manipulating tree data structures. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*. Genoa, Italy: ELRA, 2231-2234.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(b), 474-496.
- 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. (2014). *Computational methods for corpus annotation and analysis*. Dordrecht: Springer.
- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2009). Linguistic features of writing quality. *Written Communication*, 27(1), 57-86.
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in SLA: The case of

- complexity. *Applied Linguistics*, 30, 555-578.
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492-518.
- Qian, J., & Krugly-Smolka, E. (2008). Chinese graduate students' experiences with writing a literature review. *TESL Canada Journal*, 26(1), 68-86.
- Quirk, R., Greenbaum, S., Leech, G. N., & Svartvik, J. (1985). *A Comprehensive grammar of the English language*. London: Longman.
- Ravid, D., & Tolchinsky, L. (2002). Developing linguistic literacy: A comprehensive model. *Journal of Child Language*, 29(02), 417-447.
- Richards, J. C., & Schmidt, R. (2002). *Longman dictionary of applied linguistics and language teaching*. Harlow, UK: Longman.
- Schleppegrell, M. J. (1996). Conjunction in spoken English and ESL writing. *Applied Linguistics*, 17(3), 271-285.
- Scott, M. (2008). WordSmith Tools (Version 5.0) [Computer Software]. Liverpool, UK: Lexical Analysis Software.
- Sun, Y. (2008). Citation problems of Chinese ma theses and pedagogical implications. *The Journal of Asia TEFL*, 5(1), 1-27.
- Tannen, D. (1982). *Spoken and written language: Exploring orality and literacy*. Norwood, NJ: ABLEX Publishing Corporation.
- Uzuner, S. (2008). Multilingual scholars' participation in core/global academic communities: A literature review. *Journal of English for Academic Purposes*, 7(b), 250-263.
- Vande Kopple, W. J. (1994). Some characteristics and functions of grammatical subjects in scientific discourse. *Written Communication*, 11(4), 534-564.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu, HI: Second Language Teaching & Curriculum Center, University of Hawaii.
- Yeh, C. C. (2010). New graduate students' perspectives on research writing in English: A case study in Taiwan. *Journal of Academic Language and Learning*, 4(1), A1-A12.