Structural Compression in Academic Writing: An English-Chinese Comparison Study of Complex Noun Phrases in Research Article Abstracts

Zhoulin Ruan

PII: S1475-1585(18)30084-5
DOI: 10.1016/j.jeap.2018.09.001
Reference: JEAP 689

To appear in: Journal of English for Academic Purposes

Received Date: 21 February 2018
Revised Date: 29 August 2018
Accepted Date: 3 September 2018


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Title page

Title: Structural Compression in Academic Writing: An English-Chinese Comparison
Study of Complex Noun Phrases in Research Article Abstracts

Author name and affiliation:
Zhoulin RUAN (PhD)
Department of English
Xi’an Jiaotong-Liverpool University

Address:
No 111 Ren’ai Road, Dushu Lake Higher Education Town
Suzhou Industrial Park, Suzhou
Jiangsu
China, 215123
Tele: 0086 512 88161306
Email: zhoulin.ruan@xjtlu.edu.cn
Acknowledgements:

I would like to express my gratitude to the three anonymous reviewers for their insightful comments on an earlier version of this paper.

Funding: This work was supported by Research Development Fund of Xi’an Jiaotong Liverpool University and Jiangsu Ministry of Education Philosophy and Social Sciences Funding Scheme.
Structural Compression in Academic Writing: An English-Chinese Comparison Study of Complex Noun Phrases in Research Article Abstracts

1. Introduction

Over the past 30 years, research article abstracts have attracted considerable interest in the study of academic discourse. Earlier work adopted the genre analysis approach, initially developed by Swales (1990), to identify the structural moves of the texts, seeking to understand the rhetorical features of English abstracts (e.g., Bhatia, 1993; Martin, 2003; Santos, 1996; Swales, 1990). This line of research has revealed a four-part IMRD rhetorical structure (Introduction, Methods, Results, Discussion/Conclusion) that represents a summary of the different sections of research articles, though structural variation has also been found across disciplinary fields (e.g., Lorés, 2004; Pho, 2008; Samraj, 2005; Swales and Feak, 2009). More recent work has paid much attention to the interactional dimension of research article abstracts, and, following Hyland’s framework of metadiscourse (2005), investigated how academic writers attempt to engage with readers through the use of metadiscourse markers (e.g., Gillaerts and de Velde, 2010; Hu and Cao, 2011; Khedri, Heng and Ebrahimi, 2013). Other studies took a cross-linguistic perspective to explore differences and similarities between abstracts written in English and in other languages (Martin, 2003; Van Bonn and Swales, 2007), or abstracts written in English by native or non-native writers (Cao and Xiao, 2013; Hu and Cao, 2011).

Published as encapsulations of the accompanying articles, research article (RA) abstracts play a pre-eminent role in scholarly communication within academic communities, and carry some distinguishable functions, e.g., as screening devices, as previews, and as indexing help (Huckin, 2001). However, as Swales and Feak (2009) pointed out, owing to restrictions on the word count, constructing abstracts with ‘maximum efficiency, clarity, and economy’ is a difficult task even to experienced and widely published authors; and composing this highly compressed text requires sophisticated use of linguistic and rhetorical resources. There is a growing body of research studies that examines linguistic features of RA abstracts. For example, with regard to linguistic realization of rhetorical structure, a combination of grammatical subjects, verb tense and voice can help distinguish move types in abstracts (Pho, 2008). Hyland and Tse (2005a; 2005b) revealed that evaluative that is a widely employed
structure in abstracts, and functions as an important means of providing author comment and evaluation. Other researchers have looked into the interactional and interactive markers in RA abstracts written by applied linguists (e.g., Hu and Cao, 2011; Khedri, Heng and Ebrahimi, 2013). However, there are few studies so far exploring the linguistic features that are associated with the realization of structural compression in the writing of abstracts. This is an apparent, but significant gap of our understanding of the composing of this academic prose. All academic writers, irrespective of their disciplines and L1 background, need to tactically deploy linguistic strategies to produce the highly condensed summary text. For non-native English academic writers who seek to publish their scholarship in English journals, composing a compact and attractive abstract is a particularly challenging task, as it requires an advanced repertoire of L2 rhetorical and linguistic knowledge.

As a key resource for constructing scientific texts, complex noun phrases function as the syntactic unit for lexical items to be tightly packed into the clause, resulting in the high density of information in text with relatively few words (Halliday, 1993). Recent studies have shown that academic writing is structurally compressed with phrasal modifiers embedded in noun phrases, reflecting a different kind of grammatical complexity from clausal complexity in writing. Phrasal complexity and clausal complexity are considered as the two ends along ‘a single cline of grammatical complexity’ (Biber and Gray; 2016: 62). Biber and his colleagues discovered that the type of phrasal complexity is more characteristic of academic prose, which, in fact, is much more difficult to parse than clausal complexity (Biber and Gray, 2011; 2016; Biber, Gray and Poonpon, 2011). Complex noun phrases with phrasal modifiers are pervasive in written academic discourse, and reliance on such phrasal structures is ‘the unique grammatical characteristic of advanced academic writing’ (Biber, Gray and Poonpon, 2013:192). While current discourse analytical research has raised our awareness of the rhetorical structure and interactional strategies in RA abstracts, more needs to be explored with regard to linguistic resources for composing the compact text. This paper, therefore, aims to investigate the use of complex noun phrases and the patterns of nominal phrasal features in English journal abstracts written by native and non-native academic writers. We choose to make an English-Chinese comparison primarily for the reason that there exists a contrastive difference between Chinese and English languages in the phrasal structure. In Mandarin Chinese, a fundamental principle of the word order is the operation of a modifying-modified sequence, that is, relative clauses, possessives and adjectives all precede nouns (see
Kirkpatrick and Xu, 2012 for a detailed discussion). Unlike in English, in Chinese, nouns can only take pre-modification, a distinctive linguistic feature that may cause L1 transfer in composing English abstracts. In a contrastive analysis of English abstracts by English native and Chinese writers, Cao and Xiao (2013) found that Chinese writers tended to use adjective and nominal modifiers in noun phrases more frequently for the purpose of conceptual elaboration. However, their study focused on the textual variations between English native and non-native writers’ abstracts, by extracting dimensions from a comprehensive selection of linguistic features. This line of comparative research should be extended by focusing on structural compression in native and non-native EAP writing, so as to deepen our understanding of the needs and development of advanced L2 academic writers.

2. **Complex noun phrases as an indicator of structural compression**

Complex noun phrases receive extensive treatment in the literature of English grammars (e.g., Biber, Johansson, Leech, Conrad, and Fingan, 1999; Halliday and Matthiessen, 2004; Quirk, Greenbaum, Leech, and Svartvik, 1985), and as a sensitive index of style, are highly responsive to the communicative purposes and subject matters in varying types of discourse (Quirk, et al., 1985:1352). Studies on the historical change of English academic writing (Biber and Clark, 2002; Biber and Gray, 2011; 2016) revealed that there have been shifts in modification patterns within complex noun phrases: an increase in frequency for nouns as premodifiers, as well as a major extension in the range of meaning relationships underlying noun-noun sequences. Biber and Gray (2016:126) argued that ‘the most important distinguishing characteristic of academic writing is its extremely dense use of phrasal structures (i.e., structures without verbs), especially phrases functioning as noun modifiers.’ This represents a fundamental shift in the discourse style that heavily relies on nominal structures in academic English.

Research has shown that nominal groups with premodifiers are more common than those with postmodifiers in all registers in English, and multiple premodification is an efficient means by which dense information content can be packed into as few words as possible (Biber, et al., 1999: 578–597). Furthermore, as Halliday (1993) pointed out, this type of nominal group is most difficult to process, since it consists of strings of lexical words without any grammatical words in between. Biber et al. (1999: 597) describe the length and order of premodification, based on the analysis of large-scale corpora of different registers. Among
premodified noun phrases, the vast majority (70-80%) have only a single premodifier, while over 20% have two-word or more than two-word premodification. Premodifiers usually have internal logical relationships, i.e., some words modifying other premodifiers instead of the head noun (e.g., a **globally diversified business**).

Previous research has also shown the relationship between the features of modification in noun phrases and the level of L2 academic writing. The use of noun premodifiers is an important index for distinguishing proficient L2 writers from less proficient L2 writers, who relied heavily on the use of attributive adjective premodification (Parkinson and Musgrave, 2014). It was hypothesized that in L2 writing development, noun premodifiers were acquired at a later stage than the attributive adjective premodification. Proficient and novice academic writers both preferred premodification when there was only one modifier in the nominal group, but the latter used post-modification more frequently (McCabe and Gallagher, 2009). It has been argued that phrasal structures that characterize advanced academic writing, especially complex noun phrases with phrasal modifiers, should be included as an important indicator in measuring L2 academic writing development (Biber, Gray and Poonpon, 2011; 2013).

Based on large-scale corpus investigations, Biber and Gray (2010; 2016) distinguished grammatical features associated with structural elaboration from those associated with structural compression. The researchers observed that, in contrast to conversation, which is structurally more elaborated through clausal elements, academic writing is more compressed in structure and less explicit in meaning. It is characteristic of compact academic writing to embed phrasal modifiers within noun phrases. Four types of grammatical structures were identified as devices for a compressed discourse style, and all were phrasal elements used to modify a head noun: attributive adjectives (**unusual circumstances**); noun as noun pre-modifier (**surface tension**); prepositional phrases as noun post-modifier (**the scores for male and female target students**); and appositive noun phrase as noun post-modifier (Biber and Gray, 2010; 2016). Noun phrase complexity, including co-occurring features like nouns, attributive adjectives and prepositional phrases (**of-phrases as the dominant type**), is directly relevant to informational discourse (Biber, 1988). Nominal modifiers could be ranked along a cline of compression as follows (Biber and Clark, 2002: 63):

<table>
<thead>
<tr>
<th>COMPRESSED – premodifiers</th>
<th>phrasal</th>
<th>non-finite</th>
<th>relative</th>
<th>EXPANDED – EXPRESSION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postmodifiers</td>
<td>Clauses</td>
<td>Clauses</td>
<td>Expression</td>
<td></td>
</tr>
</tbody>
</table>
Among various academic genres, research article abstracts not only function as concise summaries of the accompanying articles, but also are highly compact texts due to space requirements of journals. RA abstracts therefore are representative of compressed written academic prose, and can be considered as an ideal site for exploring how complex noun phrases are used to achieve the compact discourse style by means of modification. Cao and Xiao (2013) found that Chinese writers preferred heavy premodification by adjectives and nouns and they used more frequently modified noun phrases – ‘adjective + adjective + noun’ and ‘adjective + noun + noun’. Although this study revealed interesting patterns of noun modifications in RA abstracts by native and Chinese writers, further studies on structural compression will provide a more comprehensive picture of the way by which academic writers, whether native English writers or Chinese writers, deploy the linguistic resources to compose the highly condensed text of RA abstracts.

This paper aims to explore the patterns of structural compression in RA abstracts by: 1) examining the occurrence of different types of modification embedded in complex noun phrases; 2) looking into similarities and differences in their distribution between English native and Chinese academic writers. We chose applied linguistics as the field for the investigation based on two considerations. First, despite the abundance of cumulative studies on the genre of RA abstracts in linguistics and applied linguistics (e.g., Hu and Cao, 2011; Lorés, 2004; Pho, 2008; Van Bonn and Swales, 2007), no research so far has focused on the grammatical features contributing to its compact writing. Second, and more importantly, there are systematic patterns of variation in the use of phrasal grammatical devices across disciplines in present-day academic writing (Biber and Gray, 2016). Humanities research writing relies heavily on clausal modification, but rarely on nouns as nominal pre-modifiers, whereas social science and specialist science research writing uses nouns and phrasal modifiers with extremely high frequencies (p.123). Applied linguistics, with a broad disciplinary identity, can be viewed as a sub-discipline of humanities or social science, or both (as indicated by many applied linguistics journals that are both SSCI and A&HCI indexed). An exploration into RA abstracts in applied linguistics thus will shed new light on disciplinary variations of the compact discourse style in academic writing, and contribute to the understanding of the disciplinary identification of applied linguistics.
3. Methodology

3.1. Construction of the corpus

To carry out a contrastive study of abstracts written by English native and Chinese authors, we constructed a corpus of 200 abstracts published in four applied linguistics journals: two international journals – *Applied Linguistics* (AL) and *TESOL Quarterly* (TQ); and two Chinese journals – *Chinese Journal of Applied Linguistics* (CJAL) and *Foreign Language Teaching and Research* (FLTR). The first two, published in the UK or the USA, were selected as they are the leading journals in the field, from which abstracts of native English-speaking authors were collected. *CJAL* and *FLTR* were selected as the source journals for collecting English abstracts of Chinese authors. *CJAL* is the only applied linguistics journal published in English in Mainland China, serving as a window to research for the international community. *FLTR* is ranked domestically as the top journal in the field, and is published in Chinese, with English abstracts accompanying its articles. The decision on selecting the top-tier journals was based on the assumption that abstracts written by top-tier researchers and expert writers are regarded as the model of composing effective compact texts. Features of expert academic writing would then have valuable implications for novice and less experienced English academic writers. The four journals all provide explicit requirements on the length of abstracts on their websites: 175 words by AL, 200 words by TQ, and 150 to 200 words by FLTR and CJAL. Several procedural considerations were followed to guarantee controlled comparisons. All abstracts from the issues published between 2012 and 2015 were first gathered from the four journals. Sampling abstracts published in the recent period was to reflect the features of ‘present-day’ academic research writing (Biber and Gray, 2016). Next, only abstracts reporting on empirical research were included to control potential variation in rhetorical and linguistic patterns. Abstracts of non-empirical and theoretical review articles often have varied rhetorical organization, which may result in writers’ divergence in making linguistic choices. Finally, because the four journals publish different numbers of issues each year (4 by TQ and CJAL; 6 by AL and FLTR), as well as different numbers of articles in each issue, a random sample of 50 abstracts was selected from each journal to have comparable numbers of texts. An important consideration in this study was the identification of authors’ L1 status. To determine L1 status of Chinese authors publishing in FLTR and CJAL was more straightforward, as their Chinese names and affiliations to institutions in China were highly indicative. To determine abstracts of native English-speaking authors, names connoting Anglophone origin were first identified, and then institutional affiliations
were examined to look for their biodata as available on their institutional websites. Abstracts written by authors whose names appeared to be vague in determining native English-speaking status were not included in the corpus. Although these procedures were not perfect, we were quite confident about the distinction finally established between the two sets of abstracts. Table 1 below summarizes the two parallel sub-corpora.

Insert Table 1 here

A noticeable difference between the two sub-corpora is the length of the abstracts in the four journals, regardless of each journal’s explicit guide on the text length. On average, abstracts in AL and TQ contained nearly 40 words more than those in CJAL and FLTR. This difference in sample sizes was carefully taken into account by normalizing raw frequencies per 100 words in the quantitative analysis of complex noun phrases, as well as in the comparisons of the occurrences of modifying structures between the two sub-corpora.

3.2. Identification of noun phrase modifiers

Most quantitative corpus research employs grammatical tagging and software packages to automatically identify and extract linguistic features in an entire corpus. Though this procedure could help identifying many of nominal features in the corpus, the present study required a closer examination of various types of nominal modifiers attached to every complex noun phrase in each individual text. Coding and calculating the number of modifiers per noun phrase is an important measure of phrasal complexity, but could not be automatically done by corpus tools. In addition, automatic computer programs have limited success in distinguishing prepositional phrases functioning as postmodifiers from those as adverbials (Biber and Gray, 2016: 65). This distinction is important in this study owing to the pervasive occurrences of prepositional phrases in academic prose. To accurately pinpoint how the compressed style of abstracts is realized by means of complex noun phrases, the study therefore took a qualitative approach to the identification of various types of pre- and post-modifiers in each text of the corpus. The text data was manually analysed to identify noun phrases that contain grammatical features associated with structural compression. Noun
phrases with at least one pre-modifying element or a post-modifying prepositional phrase were considered as complex noun phrases. Simple noun phrases were also identified to compare their distribution with that of complex noun phrases in the corpus. Simple noun phrases consist of determiner plus head noun (DN), e.g., *the article*, or a single noun (N), e.g. *views*. Unlike complex noun phrases, simple noun phrases are often used to establish cohesive links by means of reference (e.g., *this study*) or lexical repetition (e.g., *learners*), thus as textual organization devices, rather than devices for composing compact texts. Prepositional phrases were further determined as to whether they functioned as noun postmodifiers or as adverbials, and only those postmodifiers were included in the analysis. To illustrate, the sentence below has two prepositional phrases (Example 1): ‘about language teaching’ and ‘from their own experiences’. The former serves as the postmodifier of the head noun ‘insights’ whereas the latter functions as an adverbial related to the verbal phrase ‘gain’ (*gain… from*). Postmodifying prepositional phrases allow dense packaging of referential information in a text, though they are less compact, and more explicit about the logical relationships involved, than multiple premodifiers (Biber, et al., 1999: 607). However, compared with clausal postmodifiers, prepositional phrases are more compact, and can have complex hierarchical relations within its structure. An important point here is that these grammatical features can be treated as a cline rather than a simple dichotomy between clausal versus phrasal devices, and all phrasal devices, including post-modifying prepositional phrases, are structurally more compressed than clausal devices (Biber and Gray, 2016: 207).

Example 1:

All the plurilinguals were found to have gained *useful insights about language teaching from their own experiences*.

Premodifiers in each complex noun phrase were coded according to the grammatical category of the components, e.g., attributive adjectives (*significant effects*), noun as premodifiers (*language use*), participial adjectives (*controlled experiments*), and hyphenated compounds (*native-like proficiency*), etc.. In this study, compounds were operationalized as premodifiers that used a hyphen. Compounds can take various shapes, e.g., adjective-adjective compounds (*a cross-sectional investigation*), adjective-noun compounds (*a multiple-factor account*), or involve participial forms, e.g., noun-participle compounds (*word-based sentence production*), adjective-participle compounds (*far-reaching implications*). A special type of premodification was the sequence of adverb plus adjective/participle plus noun (aA/PN), in
which the adverb sub-modifies the adjective or participle in premodification. This sub-modification can be realized by a common adverb (a relatively homogenous sample) or markers of comparatives (the less proficient learners) or superlatives (the most used strategies). Post-modifying features, including finite relative clauses, non-finite To-clauses, and participial clauses, were not considered in the identification and analysis of complex noun phrases. This is an important decision, because, as discussed earlier, these features contribute to an elaborated discourse style, rather than the production of compressed texts. Table 2 presents the coding scheme that derived from preliminary categorization of the premodifiers and prepositional postmodifiers in the corpus. Both type and length were considered in the analysis. For example, the attributive adjective type was coded as AN and AAN, depending on whether there was one adjective or two, and noun premodifiers were distinguished between NN and NNN, according to the length of the sequence.

Insert Table 2 here

The two examples below illustrate the analysis of complex noun phrases and their modifiers in the corpus. Example 2 was an opening sentence of a Chinese writer’s abstract. In addition to a simple noun phrase (the factors), it contained eight complex noun phrases (underlined), with five different types of pre-modifiers (annotated in italics). In Example 3, an opening sentence of a native English writer’s abstract, one simple noun phrase (this study) and five complex noun phrases were present, the latter of which contained five different constructions of premodifiers and three prepositional postmodifiers.

Example 2 Chinese writer’s abstract

The present study AN aimed to investigate the factors DN associated with English word reading NNN in Mandarin-speaking children CN by administering an English word reading test NNN, a phonological awareness test ANN, the Woodcock word attack NNN and basic cognitive tests AAN in seventy-six fourth and fifth graders AAN.
Example 3 Native English writer’s abstract

This study provides an empirically based contribution to the growing body of research using conversation analysis as a methodical tool for analyzing functions of action types through interaction in a classroom setting.

Once the coding of all the abstracts was completed, an independent coding was undertaken by a trained applied linguist. Two sets of modifiers identified by the author and the independent coder were then compared in each abstract, and an average agreement of 87% was reached. Most of the disagreed cases were resolved after further discussions between the coders, which resulted in a final inter-coder agreement of 96%. The remaining problematic cases, mainly concerning prepositional phrases as modifier or adverbial, were excluded in further analysis. After identifying these various types of modification in the entire corpus, we counted and compared the frequency of each type of premodification between the two sub-corpora. In taking into account the length variation in the two sets of abstracts, raw frequencies were normalized to a common base of 100 words, owing to the fact that the abstracts in the corpus typically contained less than 200 words.

3.3 Statistical measures used in the study

We first conducted t-tests as a method to look for differences between English native and Chinese writers’ abstracts. As multiple comparisons were performed on the same sets of data, Bonferroni’s Correction was adopted to adjust the alpha level for robust analyses, which was set at p < 0.004 in reporting significant differences. All t-test results were reported with the accompany of the effect sizes calculated by means of the Cohen’s d. Cohen’s d is a descriptive statistic that expresses the mean difference between groups beyond finding significance, and provides an estimate of the magnitude of the effect in question (Plonsky, 2015). As general benchmarks, a d value of 0.5 would be interpreted as a medium effect and 1.0 as a large effect in this study.
4. Results

We first present the distribution of simple and complex noun phrases, as well as the distribution of different types of modifiers in complex noun phrases in the two sub-corpora. This is followed by comparisons of their occurrences through statistical analyses. We then take a closer look at some structural components of the modifiers, in order to uncover their differences in the two sets of abstracts written by English native and Chinese writers.

4.1 Distribution of noun phrases and each type of modifier between sub-corpora

We first compared the distribution of simple and complex noun phrases in the two sub-corpora. As shown in Table 3, a noticeable difference was the occurring frequencies of simple noun phrases, with native English writers using nearly twice more than Chinese writers (1055 vs 643). In addition, the ratio of simple noun phrases to complex noun phrases was 0.35 for Chinese writers (643/1837) and 0.53 for native English writers (1055/1994). After normalizing the raw frequencies per 100 words in each text, the average number of complex noun phrases was 11.6 in native English writers’ abstracts and 13.8 in Chinese writers’ abstracts, and the number of simple noun phrases was 6.1 and 4.9 respectively. Independent sample $t$-tests were carried out to check whether the differences between two sub-corpora were statistically significant ($p < 0.004$, the alpha was set at this level owing to multiple comparisons on the same data). The results show that complex noun phrases were significantly more in Chinese writers’ abstracts than in native English writers’ abstracts ($p<0.000$). On the other hand, native English writers used significantly more simple noun phrases than Chinese writers in composing abstracts ($p<0.001$). These differences were further confirmed by means of the Cohen’s $d$ for effect sizes, whose values were 1.27 and 0.51 respectively, meaning a large to medium effect (Plonsky, 2015). The results suggest “real” differences in the use of simple and complex noun phrases between native English writers and Chinese writers.

Insert Table 3 here
Apart from the occurring frequencies of complex noun phrases, another important indicator of the compressed writing style is the average number of the modifying features per noun phrase. The more pre- and post-modification attached to a noun phrase, the more complex its internal structure, which represents a more compressed writing style. For example, in the sentence (Example 4) below, within the noun phrase with ‘significance’ as the head noun, there are two attributive adjectives (AN), i.e., ‘cultural’ and ‘linguistic’, a compound premodifier (CN), i.e., ‘cross-racial’, and two postmodifying of phrases (PrepOF), whereas the noun phrase with ‘discourse’ as the head noun is pre-modified by one single sequence of construction: an attributive adjective and a nominal modifier (ANN). The first noun phrase therefore represents a more compact phrasal structure than the second one.

Example 4

This article examines the cultural significance AN of PrepOF cross-racial embodiments CN of PrepOF linguistic signs AN that may be legible as ‘black’ within mainstream US discourse ANN.

Table 4 below shows the comparison of the number of modifiers per noun phrase between two sub-corpora. On average, there were more pre- and post-modifications attached to noun phrases in native English writers’ abstracts (1.48) than in Chinese writers’ abstracts (1.35). This difference was found statistically significant (p < 0.000), which was further confirmed by the Cohen’s d value of 0.55 for the effect size, suggesting medium strength in effect. The finding indicates a contrasting, but interesting pattern of compressed writing between English native and Chinese academic authors. Whilst the higher number of complex noun phrases used by Chinese authors suggests they relied on a more frequent use of complex noun phrases as a way of packaging information in English abstracts, native English writers preferred to embed more modifying constructions within noun phrases, thus increasing the complexity of the phrasal structure, to convey information.

Insert Table 4 here
Table 5 shows raw frequencies, normed frequencies, and percentages of each type of premodifier and prepositional phrase that were derived from the analysis of the two sub-corpora. The raw frequencies were normalized per 100 words for the purpose of reliable comparison. A close look at the distribution of each type of modifier in the table shows similarities and differences between native English writers’ abstracts and Chinese writers’ abstracts. In both sub-corpora, the top three types of modifiers consisted of attributive adjectives (AN), the postmodifying prepositional of phrase, and noun as noun premodifier (NN), and they amounted to more than 60% of the total number of modifiers. But their ranking orders were different between two sub-corpora. In native English writers’ abstracts, the most frequent type was the prepositional of phrase (24.5%), followed by attributive adjective (22.1%), and nominal premodifier (15.3%), whereas in Chinese writers’ abstracts, it was attributive adjective (24%), followed by of phrase (19.9%) and nominal premodifier (18.6%). Except for these three types, the other premodifiers and prepositional phrases had low distributions in both sub-corpora, within the range between 4% and 8%. Though a small proportion of less than 2% in each sub-corpus, the structural sequence ‘adverb + adjective/participle + noun’ was a distinctive linguistic device for producing compressed texts. Using adverbs to sub-modify attributive adjectives or participles not only rendered another level of modification in the noun phrase, thus increasing the complexity of its internal logical structure, but also semantically allowed the writer to compress additional information into premodification.

Insert Table 5 here

Table 6 below shows the comparisons of the normalized occurrences of each type of modifier in the two sub-corpora. Mean occurrences of all ten types of pre- and post-modifications were compared by employing t-tests for statistical significance (p<0.004), and the results were further checked by Cohen’s d for effect sizes. Of the ten types of structural compression, noun premodifier (NN) and adjective plus noun premodifier (ANN) were found to have statistically significant differences between English native and Chinese writers, which were further confirmed by Cohen’s d values of 0.43 and 0.57 respectively. Along with the occurrence of multiple noun sequence (NNN), an emerging pattern is that Chinese writers used more nominal premodifiers than native English writers. Chinese writers also tended to
use the construction of adjectives plus head noun (AN) more often, though the difference was not statistically significant. On the other hand, the occurrence of post-modifying of phrases, though abundant in both sub-corpora, had an opposite trend: native English writers tended to use more postmodifying of phrases than Chinese writers. No significant difference was found between two sub-corpora in the use of the less frequent types of constructions, including adjective plus adjective (AAN), other prepositional postmodifier, compound as premodifier (CN), participial premodifier and adverb sub-modification.

Insert Table 6 here

4.2 Structural patterns of premodifications across two sub-corpora

Apart from comparing the occurring frequencies, our analysis also examined distinct structural patterns of premodifications between native and Chinese academic writers. We were particularly interested in two distinctive types – hyphenated compounds and adverb sub-modification – both of which involve additional levels of premodification, thus representing a more compressed style of writing. Though not as frequent as attributive adjectives and nominal premodifiers, as noted above, hyphenated compounds were embedded in noun phrases as compressed structures, resulting in ‘a compact and integrated expression of information’ (Biber, et al., 1999:533). They were used as important linguistic resources in the process of planning, drafting and editing to produce compact abstracts. A close examination of the construction of hyphenated compounds reveals some qualitative differences in structural formation. For Chinese academic writers, one common form of compound consists of numerals plus time-related nouns, used to express temporal meaning in premodification, for example:

a 22-week semester, 53 third-year English majors, each 90-minute session, a ten-week study, first-year English majors

Another form of compound in Chinese writers’ abstracts takes the shape of an adjective-noun combination, but is embedded, as a whole, attributively to the noun phrase. As the adjectives within the compound chunks (cross, lower, high, post) modify its following noun,
not the head of the noun phrase, this means a further structural level in premodification. For Chinese writers, such compound constructions created an additional structural slot for inserting attributive adjectives in premodification.

*cross-language* matching model, *lower-level* learners, *post-task* stage, *high-score* stage, *after-class* reflection, *high-level* lexical component skill

A third form of compound in Chinese writers’ abstracts involves a prefix being added to a noun or adjective, which can be seen as the result of lexical expansion through affixation.


In the field of applied linguistics, many compounds of these three formations are familiar enough to be seen as ‘lexicalized expressions’ (Biber et al., 1999: 535). The compounding sequence of numerals plus time-related nouns is commonly used in describing methodology in reporting empirical studies. The formation of adjective-noun compound derives from the lexicalization of two collocates, whereby the first (e.g., *high, lower*) is modifying the second (e.g., *level, score*). Compounds formed by affixation can be viewed as single lexical items, but with a closed set of prefix items, have a narrow range of internal semantic potentials. The frequent occurrences of these compounds therefore suggest that Chinese writers may draw on their lexical knowledge in their attempt to compress information into the noun phrase in academic writing. As the formations of these compounds are rather fixed, the results also indicate features of formulaic language use in academic writing by Chinese writers.

In native English writers’ abstracts, a common form of compound chunk involves the participial verb *-ed* or *-ing*. Constructions which consist of a noun plus *ed*-participle were especially abundant:


Another frequent type of compound involves a noun or adjective plus *ing*-participle:

the *fastest-growing* group, powerful *identity-affirming* purposes, *German-speaking* learners, *test-taking* strategy, *ideology-reflecting* potential, *far-reaching* implications
Unlike lexicalized compounds in Chinese writers’ abstracts, a wide range of nouns and participles were incorporated into these types of compounds in native English writers’ abstracts. Furthermore, these compounds can be expanded into full clausal expressions, often relative clauses, in an elaborated way of writing, such as, *rules that are generated by learner, the group that is growing the fastest*. Using such compounds is an efficient means by which the writers can convert lengthy clausal expressions into compact phrasal structures in abstracts. As an alternative to the elaborated, clausal postmodification, these compounds allow the writers to compress the same amount of information into the structure of premodification of noun phrases. This type of participial compound therefore is highly productive, illustrating the point that compounding can be generally viewed as “prefixation with open-class items” (Quirk, et. al., 1985:1568).

As noted above, the modifier sequence of adverb plus adjective or participle is also a distinctive type of premodification, as it contains an internal logical structure of its own. At the clause level, adverbs co-occur most commonly with verbs and function as adverbials. Within the structure of complex noun phrases, adverbs preceding adjectives or participles occur as phrasal elements, functioning as premodification of the head noun, e.g., *grammatically correct sentence, semantically unrelated pairs*. These constructions represent a logical structure of sub-modification within the complex noun phrase, as the adverb serves as the modifier of a pre-modifying element. Like compounds, they are the important resources to package dense information into complex noun phrases. A scrutiny of the structural components reveals interesting differences between native and Chinese writers. In the abstracts of Chinese writers, *more* and *most* were often used to modify attributive adjectives for the purpose of phrasal comparison. Both words functioned as phrasal markers of degree prior to gradable adjectives.

a more influential role, the most visible representation, more adequate visual teaching materials, most meaningful predicators

When other adverbs were used, they also tended to modify an adjective, and to express the writer’s stance towards propositional content (e.g., *highly, significantly*).

the seemingly same changes; comparatively few studies, a significantly higher occurrence, highly proficient Chinese-English bilinguals, grammatically correct sentence
In native English writers’ abstracts, however, the use of adverbs appeared to be semantically more varied, and to modify a participial verb in the premodification of the head noun, e.g. *socially situated, empirically based*. Like compounds involving participial forms, these compressed phrasal constructions can also be expanded into an elaborating, post-modifying relative clause of the head noun, such as *symbolic resources that are socially situated, classrooms that are ethnically and linguistically changing and challenging*.

*socially situated* symbolic resources, *increasingly sophisticated* functions, an *empirically based* contribution, a homogeneous, *culturally and linguistically united* transnational community, *ethnically and linguistically changing and challenging* classrooms

4. Discussion

The present study has explored the linguistic features that are associated with the condensed discourse style of research article abstracts written by native English-speaking authors and Chinese authors. Drawing on abstracts collected from two international journals and two Chinese journals in applied linguistics, we made a comparative analysis of the phrasal structures of complex noun phrases that were used as linguistic resources for the realization of structural compression.

Our analysis has shown significant differences in the distribution of simple and complex noun phrases in RA abstracts between native and Chinese writers. Chinese writers used more complex noun phrases in writing English abstracts, whereas native English writers used more simple noun phrases in their texts. Chinese writers’ preference to relying on a higher number of complex noun phrases is congruent with the findings reported in Cao and Xiao’s (2013) study. On the other hand, native English writers were inclined to use more modifiers embedded in noun phrases, rendering more complex phrasal structures. Apart from these distinctive distributions of noun phrases, our analysis also found that the majority of modifiers were comprised of attributive adjectives, noun modifiers, and the prepositional *of* phrases in both native and Chinese writers’ abstracts. Previous studies have also reported the prominence of these modifying features in academic writing (e.g. Biber and Gray, 2011; Parkinson and Musgrave, 2014), and in English abstracts in particular (Cao and Xiao, 2013). Furthermore, in Chinese writers’ abstracts, the occurrences of noun premodifiers and multiple
noun sequences were significantly higher, whereas in native English writers’ abstracts, there was a more frequent use of the post-modifying *of* phrases. This difference may be partially attributable to the influence of L1 transfer to L2 academic writing (Cao and Xiao, 2013), because in Chinese, the noun phrase follows a modifying-modified order, and its construction can only take premodifiers. Chinese writers therefore may retain a preference for modifiers based on L1 in English academic writing. Prepositional phrases are by far the most common type of postmodifier in all registers, and the majority of such postmodifiers are *of*-phrases (Biber, et al. 1999: 635). Furthermore, noun phrases followed by a post-modifying *of*-phrase constitute the most common lexical bundles in academic writing, and convey a wide range of functions, e.g., identifying abstract qualities, describing processes or events, marking existence or presence, or describing physical entities (Biber, et al. 1999: 1015). Chinese L2 writers may not have developed the register awareness of the wide range of functions that prepositional *of* phrases are able to perform in academic writing, and as a result, they made less frequent use of this postmodifying feature that is ubiquitous in native speakers’ academic writing.

An interesting finding of this study is concerned with the structural patterns of hyphenated compounds and sub-modification, both of which are distinctive grammatical devices for compressing dense information in noun phrases. Formations of compounds were found to be distinct between native and Chinese writers’ abstracts. Chinese writers tended to add a modifying adjective to a noun or use lexicalized expressions to form compounds. Compounds consisting of adjective and noun created an extra structural slot for inserting attributive adjectives in the premodification of noun phrases. Another salient feature in Chinese writers’ abstracts is the common use of prefixes to form compounds, and as a result of lexicalization, such compounds are often widely accepted as established lexical items. As prefixes are the closed set of items, this type of compound is less productive than formations comprised of open-class words. Chinese writers also used comparative and superlative adverbs – *more* and *most* – more often as phrasal markers of degree to modify gradable adjectives. By contrast, noun-participle compounds are a more common type of premodifying construction in native English writers’ abstracts. This type of compound is also more productive as the two components draw on open lexical classes with a range of potential semantic relations. Research has shown that using noun-participle compounds as premodifiers has notably increased in present-day English academic writing (Biber and Gray, 2016). In addition, sub-modification in native English writers’ abstracts often involved the participial verb *-ed* or –
ing, the grammatical feature that can be expanded into a clausal expression in an elaborated writing style. These findings suggest that in packaging dense information in complex noun phrases, Chinese writers tended to rely on their lexical knowledge in producing compressed structures, whilst native English writers drew on their repertoire of syntactic knowledge to convert clausal structures to phrasal premodifications. Native English writers also used more productive types of compounding and sub-modification in premodification than Chinese writers who drew on more lexicalized expressions. Hu and Cao (2011) found that Chinese writers used different metadiscourse markers in academic article abstracts from native English writers, which was partly attributable to "culturally preferred rhetorical strategies". The findings of this study lead us to argue that there are also notable distinctions in the employment of linguistic strategies for structural compression in advanced academic writing between native and Chinese writers. Such distinctions can be accounted for in terms of Chinese L1 influence and the availability of the faculty for sophistication in using English as a second language in advanced academic writing.

The findings of the study have important implications for genre analysis of research article abstracts in applied linguistics as well as English academic writing. Much of previous research on RA abstracts has focused on the identification of the rhetorical structure (e.g., Martin, 2003; Swales, 1990) and the examination of metadiscourse features (e.g., Gillaerts and de Velde, 2010; Hu and Cao, 2011). However, we are still short of understanding of grammatical features for achieving the compact discourse style in RA abstracts. The findings of the study explain the abundance of complex noun phrases in RA abstracts by both native and non-native academic writers in applied linguistics. The use of complex noun phrases is motivated by the drive for 'economy of expression' (Biber and Gray, 2016), that is, conveying the maximum amount of information in the fewest words possible. Though the study drew on the analysis of abstracts published in applied linguistics journals, structural compression is the functional motivation shared by abstracts in academic journals of all disciplines. The patterns of complex noun phrases discovered in this study, therefore, are revealing about the generic features of research article abstracts as a sub-genre. Furthermore, as pointed out earlier that applied linguistics has a broad discipline identity, the high frequencies of noun and phrasal modifiers bear resemblance to the patterns of modifiers in social science research writing, but are distinct from humanities research writing (Biber and Gray, 2016). This leads us to suggest that it may be more plausible to view applied linguistics as a sub-discipline of social sciences rather than humanities.
Pursuing economy of expression can result in fewer words in academic writing, but, as compressed structure is implicit in terms of relations between components, the meaning becomes less explicit. For example, with no function words in-between, the internal logical structure of noun sequences is not explicit, thus complex to process the meaning (Biber et al., 1999). This increases the difficulty of comprehension as readers have to infer the intended logical relations to figure out the highly dense packaging of referential information. Though a typical feature of Chinese writers’ abstracts, it has been suggested by many Chinese journals that long sequences of adjective and nominal premodifiers be avoided in English abstracts, in calling for a shift of dense information writing to a more readable style with simpler structures (Cao and Xiao, 2013). In academic writing, therefore, there is a tension between maximum economy of expression and clarity of meaning, and it is especially the case in the writing of journal article abstracts by L2 academic writers. Highly condensed structures in premodification may result in the opaqueness of meaning, and affect the readability of the abstract, which is opposite to its primary communicative function. We would argue that it is of particular importance to keep a balance between structural compression and clarity of meaning in academic writing, in particular in the writing of research article abstracts.

6. Conclusion

This study aimed to fill in the gap of research on grammatical features that account for the compressed discourse style in native and non-native academic writing in applied linguistics. The study reveals distinct linguistic strategies used by both groups of writers in achieving structural compression. The findings cast new light on academic writers’ use of complex noun phrases to fulfil the challenging task of composing a compact summary text. We would argue that explicit awareness of compact grammatical features is crucial to L2 academic writers in achieving their rhetorical and pragmatic goals in academic writing. The study has a few limitations that need to be addressed in future research. First, this study has drawn on a small corpus of research article abstracts from one discipline, which may have skewed our understanding of the breadth of academic writing. The findings need to be interpreted tentatively and it is not our attempt to make broad generalizations about academic writing. Second, though the data was coded meticulously, and a confident level of inter-coder agreement was reached, automatic extraction by corpus tools would reduce potential issues arising from manual coding. Tagging the raw texts could facilitate the identification of
different types of modifications in the entire corpus. Finally, although we explored ten categories of grammatical features associated with structural compression, the scheme could be further refined. For example, a more fine-grained analysis of hyphenated compounds could further elaborate its internal constructions with a mixture of components. Future research, therefore, can be undertaken by looking into larger corpora of research article abstracts gathered from a wide range of the disciplinary areas. This will enable us to explore the patterns of noun phrase modifications in abstracts in comparison with other genres or subgenres of academic writing, e.g., their corresponding research articles, to reveal the density of information in various academic writing. As noun phrases are linguistic resources typically for conveying technical contents in scientific writing, further investigations can look into the discipline-specific features of structural compression in academic writing. More research along these lines will contribute to our knowledge of the linguistic features characteristic to compact discourse styles, which can be usefully drawn upon for the instruction of advanced academic writing to serve effective scholarly communication.
References:


Table 1 Description of the corpus of abstracts

<table>
<thead>
<tr>
<th>Sub-corpus</th>
<th>No of texts</th>
<th>No of words</th>
<th>Mean length</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS writers’ abstracts</td>
<td>100</td>
<td>17,166</td>
<td>171.7</td>
<td>23.2</td>
</tr>
<tr>
<td>AL</td>
<td>50</td>
<td>8,183</td>
<td>163.7</td>
<td>13.5</td>
</tr>
<tr>
<td>TQ</td>
<td>50</td>
<td>8,983</td>
<td>179.7</td>
<td>28.4</td>
</tr>
<tr>
<td>Chinese writers’ abstracts</td>
<td>100</td>
<td>13,214</td>
<td>132.1</td>
<td>29.3</td>
</tr>
<tr>
<td>CJAL</td>
<td>50</td>
<td>6,496</td>
<td>129.9</td>
<td>29.6</td>
</tr>
<tr>
<td>FLTR</td>
<td>50</td>
<td>6,718</td>
<td>134.4</td>
<td>29.0</td>
</tr>
</tbody>
</table>

Table 2 Grammatical features associated with structural compression

<table>
<thead>
<tr>
<th>Grammatical feature</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attributive adjectives + Noun (AN)</td>
<td><em>the explicit intervention, a new approach</em></td>
</tr>
<tr>
<td>Noun + Noun (NN)</td>
<td><em>vocabulary needs, research participants</em></td>
</tr>
<tr>
<td>Adjective + adjective + Noun (AAN)</td>
<td><em>active social agents, explicit cognitive processes</em></td>
</tr>
<tr>
<td>Adjective + Noun + Noun (ANN)</td>
<td><em>oral production data, a regional accent recognition instrument</em></td>
</tr>
<tr>
<td>Of phrase as noun postmodifier (PrepOF)</td>
<td><em>the use of metaphor, the limitations of traditional research</em></td>
</tr>
<tr>
<td>Other prepositional phrase as noun postmodifier (PrepTO, PrepIN, etc.)</td>
<td><em>our understanding of variation in academic vocabulary, a significant effect on use of the passive voice</em></td>
</tr>
<tr>
<td>Noun + Noun + Noun (NNN)</td>
<td><em>grammaticality judgement task, discourse community membership</em></td>
</tr>
<tr>
<td>Compounds + Noun (CN)</td>
<td><em>the small-scale study, corpus-driven analysis</em></td>
</tr>
<tr>
<td>Participle + Noun (PN)</td>
<td><em>the targeted forms, a recurring theme</em></td>
</tr>
<tr>
<td>Adverb + Adjective/Participle + Noun (aA/PN)</td>
<td><em>increasingly sophisticated functions, a relatively homogeneous sample</em></td>
</tr>
</tbody>
</table>

Table 3 Occurrences of simple and complex noun phrases between two sub-corpora

<table>
<thead>
<tr>
<th></th>
<th>Chinese writers’ abstracts (n=100)</th>
<th>Native writers’ abstracts (n=100)</th>
<th>Mean difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw freq.</td>
<td>Mean (per 100 words)</td>
<td>SD</td>
</tr>
<tr>
<td>Simple NP</td>
<td>643</td>
<td>4.9</td>
<td>2.12</td>
</tr>
<tr>
<td>Complex NP</td>
<td>1837</td>
<td>13.8</td>
<td>1.79</td>
</tr>
</tbody>
</table>
Table 4 Average number of modifiers per noun phrase between sub-corpora

<table>
<thead>
<tr>
<th></th>
<th>Mean (per NP)</th>
<th>SD</th>
<th>p (t-test)</th>
<th>Cohen’s $d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese writers’ abstracts (n=100)</td>
<td>1.35</td>
<td>0.21</td>
<td>0.000*</td>
<td>0.55</td>
</tr>
<tr>
<td>Native writers’ abstracts (n=100)</td>
<td>1.48</td>
<td>0.26</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5 Frequencies and proportions of each type of modifier in two sub-corpora

<table>
<thead>
<tr>
<th>Modifiers</th>
<th>Abstracts of Chinese writers</th>
<th>Abstracts of native English writers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw freq.</td>
<td>Per 100 words</td>
</tr>
<tr>
<td>AN</td>
<td>595</td>
<td>4.50</td>
</tr>
<tr>
<td>Of phrase</td>
<td>493</td>
<td>3.73</td>
</tr>
<tr>
<td>NN</td>
<td>459</td>
<td>3.47</td>
</tr>
<tr>
<td>ANN</td>
<td>217</td>
<td>1.64</td>
</tr>
<tr>
<td>Other prep phrase</td>
<td>176</td>
<td>1.30</td>
</tr>
<tr>
<td>NNN</td>
<td>166</td>
<td>1.27</td>
</tr>
<tr>
<td>CN</td>
<td>121</td>
<td>0.96</td>
</tr>
<tr>
<td>PN</td>
<td>107</td>
<td>0.83</td>
</tr>
<tr>
<td>AAN</td>
<td>107</td>
<td>0.83</td>
</tr>
<tr>
<td>aA/PN</td>
<td>35</td>
<td>0.28</td>
</tr>
<tr>
<td>Total</td>
<td>2476</td>
<td>18.78</td>
</tr>
</tbody>
</table>
Table 6 Comparisons of occurrences of modifiers after normalization (per 100 words)

<table>
<thead>
<tr>
<th></th>
<th>Chinese writers’ abstracts (n=100)</th>
<th>Native writers’ abstracts (n=100)</th>
<th>Mean difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>AN</td>
<td>4.50</td>
<td>2.46</td>
<td>3.86</td>
</tr>
<tr>
<td>Of phrase</td>
<td>3.73</td>
<td>1.86</td>
<td>4.27</td>
</tr>
<tr>
<td>NN</td>
<td>3.47</td>
<td>2.34</td>
<td>2.63</td>
</tr>
<tr>
<td>ANN</td>
<td>1.64</td>
<td>1.46</td>
<td>0.97</td>
</tr>
<tr>
<td>Other prep phrase</td>
<td>1.30</td>
<td>1.27</td>
<td>1.39</td>
</tr>
<tr>
<td>NNN</td>
<td>1.27</td>
<td>1.19</td>
<td>0.88</td>
</tr>
<tr>
<td>CN</td>
<td>0.96</td>
<td>0.96</td>
<td>1.07</td>
</tr>
<tr>
<td>PN</td>
<td>0.83</td>
<td>0.79</td>
<td>1.10</td>
</tr>
<tr>
<td>AAN</td>
<td>0.81</td>
<td>0.98</td>
<td>0.85</td>
</tr>
<tr>
<td>aA/PN</td>
<td>0.28</td>
<td>0.50</td>
<td>0.31</td>
</tr>
</tbody>
</table>
Author Bio

Author’s bionote

Zhoulin RUAN is Professor of Applied Linguistics at the Department of English at Xi’an Jiaotong Liverpool University. He received his PhD in applied linguistics from the University of Reading, UK. His research interests include academic writing, metacognition and self-regulated language learning, EAP/ESP in the Chinese context, and Systemic Functional Grammar.
学霸图书馆
www.xuebalib.com

本文献由“学霸图书馆-文献云下载”收集自网络，仅供学习交流使用。

学霸图书馆（www.xuebalib.com）是一个“整合众多图书馆数据库资源，提供一站式文献检索和下载服务”的24小时在线不限IP图书馆。

图书馆致力于便利、促进学习与科研，提供最强文献下载服务。

图书馆导航：
图书馆首页 文献云下载 图书馆入口 外文数据库大全 疑难文献辅助工具