A Corpus-Driven and Functional Analysis of the Use of NP-*of* Phrases in English Thesis Abstracts by Novice and More Advanced Writers

Jing Li

Abstract Structural complexity, a recognized feature of academic writing, has been recently proven by Biber and Gray (Journal of English for Academic Purposes, 9: 2–20, 2010) to be embedded in nominal expressions with phrasal modifiers. However, little has been done to identify the problems that novice Chinese academic writers might face when using nominal expressions to compress information, especially from a functional perspective. This chapter explores the most common phrasal modifier—*of*–phrases—aiming to find how novice Chinese academic writers in computer science differ from more advanced science writers regarding the discourse-semantic function in which nominal expressions with *of*-phrases as postmodifiers (NP-*of* phrases) perform. NP-*of* phrases retrieved from two self-built corpora of master's thesis abstracts from China National Knowledge Infrastructure (CNKI) and ProQuest were investigated both quantitatively and qualitatively. Our findings show that novice Chinese academic writers use NP-*of* phrases more frequently but do not compress information as effectively as more advanced writers. For the semantic function, they tend to underuse NP-*of* phrases with an internal nominalized relationship and overuse some head nouns before *of*-phrases in certain abstract moves. For the discoursal function, they use fewer NP-*of* phrases in the verbalized cause–effect relationship, especially in the problemizing step, the result move and conclusion move. These deficiencies may hinder novice Chinese academic writers' ability to produce compact and compelling academic discourse with the meaning-making function of the NP-*of* phrase enabled by its structural complexity.

Keywords Structural complexity · Academic writing · *Of*-phrases · Discourse-semantic function · Cause–effect relationship

J. Li (✉)
School of Humanities, Beijing University of Posts and Telecommunications (BUPT), Beijing, China

1   Introduction

Numerous studies have indicated that academic writing is more structurally complex than speech, measured by the mean length of "*t*-units," and the number of subordinate clauses per unit (e.g., Brown & Yule, 1983; Chafe, 1982; Kroll, 1977; O'Donnell, 1974). Supported by a large-scale corpus investigation, Biber and Gray have challenged this stereotyped notion of complexity in academic writing by redefining it as "being structurally compressed, with phrasal (non-clausal) modifiers embedded in noun phrases" (3). As a matter of fact, a range of other studies have identified the nominally oriented language as a key feature in academic writing (e.g., Halliday, 1988; Ravelli, 1996; Hunston, 2002; Hyland, 2009; Coffin, 2009), though none of them have explicitly associated it with the interpretation of complexity in academic writing.

When they compared the complexity feature of academic writing to that of conversation, Biber and Gray have discovered that "the most striking difference from conversation is the use of prepositional phrases as noun postmodifiers. Many of these are *of*-phrases (e.g., the *participant perspective of members of a lifeworld*)" (7). There is also evidence that the nominal structures in academic writing represent a high degree of complexity and will be acquired in later developmental stages of learners (e.g., Biber, Gray, & Poonpon, 2011; Lu, 2011; Parkinson & Musgrave, 2014). Therefore, for English for Academic Purposes (EAP) instructors, it should be an important part of their teaching syllabi to train novice academic writers to use nominal phrases, especially the complex ones, such as NP-*of* phrases to create a compressed academic style.

Postgraduates in Computer Science at Beijing University of Posts and Telecommunications (BUPT) have never taken basic EAP courses for decades, let alone those courses which aim to help learners develop a compact style in English academic writing. To prepare such a course for upcoming students enrolled in the year of 2018 to fill the gap, we conducted a contrastive study to compare the English master's thesis abstracts of our students between 2010 and 2015 from China National Knowledge Infrastructure (CNKI)—the universally available research product in school for most students before they take a real academic path—with abstracts between 2000 and 2015 from ProQuest written by postgraduates with similar academic background from universities in English-speaking countries, mostly in the USA. Since we tried to match the two corpora for topics, we set a broader time range for our search in ProQuest so that we can increase our chances of success in finding the best match in ProQuest for the selected CNKI abstracts. Though there might be some individual variations within each group, we assume that the CNKI abstracts represent the novice writer level and the ProQuest abstracts represent a more advanced level of proficiency, because the former samples were written by students who had never taken any EAP training while the latter were written by students likely to have been assisted by their universities in the USA which generally offer EAP courses and editing services from writing centers. Our aim is to inform future EAP writing instruction in China of our new findings by

identifying the difficulties novice Chinese academic writers might face in using NP-*of* phrases to develop complexity in academic writing. In particular, we focus on the genre of English-language abstracts which require a highly informational and condensed writing style. This study is also part of widespread efforts to cultivate EAP learners' rhetoric awareness of the prominent grammatical structures in academic writing.

2   NP-*of* Phrase and the Semantic Approach to Its Complexity in Academic Writing

Though Biber and Gray attribute the complexity in academic writing to the high frequency of nominal expressions, they mainly focus on the grammatical aspect of complexity—the structure of "multiple levels of phrasal embeddings" (19) without exploring how the linguistic form of nominal expressions should function to meet the meaning-based, contextual, and communicative demands of complex academic discourse. According to Systemic Functional Linguistics (SFL) grammar, "grammar is seen as a resource for making meaning—it is a 'semanticky' kind of grammar" (Halliday 31). A functional analysis is therefore needed to guarantee that novice academic writers will not only learn the linguistic forms of complexity but also know how and when to use them to generate proper meaning in line with the requirement of specific writing tasks. This element is still missing in a number of recent studies of linguistic complexity (e.g., Yoon, 2017; Parkinson & Musgrave, 2014). However, there has been recent growing interest in the discourse-semantic role the construct of complexity may play (e.g., Ryshina-Pankova, 2015; Housen Kuiken & Vedder, 2012). Housen, Kuiken, and Vedder define linguistic complexity in L2 writing as "formal and semantic-functional properties" (4). Ryshina-Pankova argues that "the majority of research on linguistic complexity have favored the formal properties of complexity and operationalized this construct as grammatical and lexical complexity (Bulte and Housen 23). What appears to be missing from the research is a focus on the discourse-semantic aspects of linguistic complexity defined as a focus on the link between the complexity of linguistic forms and its function in construction of meaning (semantics) in texts (discourse) at different levels of L2 abilities, an area of investigation that should be of particular interest to L2 writing research and instruction" (52).

With regard to the discourse-semantic function of nominal phrases, Biber and Gray briefly mentioned: "they are more economical, and they allow faster reading by experienced knowledgeable readers" (11). Halliday and Martin also discuss how a nominal rather than clausal expression of meaning makes text denser (76). Carroll, a once frustrated history teacher in college found that nominalization could help his students build stronger arguments in their writing (24). Ryshina-Pankova also argues that nominalization can enable a better understanding with "the abstracting and theorizing" needed to "fulfill the communicative purposes of particular situations" (55).

The studies outlined in the preceding paragraph all reflect a view that the desirable use of complex nominal expressions can realize a discourse-semantic function of facilitating information flow and advancing writers' arguments efficiently in academic discourse. However, how to understand and evaluate this meaning-based function of nominal phrases, specifically complex ones, such as NP-*of* phrases remains an unsolved problem. This lack of knowledge will make it impossible for EAP instructors to track learners' progression in acquiring the complexity feature of academic writing. Focusing only on the literature of NP-*of* phrases, we still cannot find a satisfying answer.

One focus of studies concerning the NP-*of* phrase is its diverse meaning relationships between the head noun and its *of*-postmodifier (Biber & Gray, 2010; Cai, 2003; Ping & Yan, 2012; Sinclair, 2005; Quirk, Greenbaum, Leech & Svartvik, 1985), because "Prepositional phrases beginning with 'of' are the most numerous postmodifiers and convey very rich meaning relationships with the head noun" (Sinclair 129). Most of these studies elaborate on different semantic relationships within NP-*of* phrases either with or without a comparison with the s-genitives. Only a few researchers have explicitly mentioned the effect this structure may produce, such as Cai (2003), who mentions that the appositive relationship between the *of*-postmodier and the head noun most effectively compress information of a clause (39) as well as Li Ping and Cao Yan who studied the condensing function of *of*-postmodifiers in research paper titles based on the titles' semantic structures (322–324). Biber and Gray have also explored the various semantic relationships within the nominal phrases including the ones with *of*-postmodifiers to demonstrate the implicit style in academic writing (7–12). Despite the differences in the research motives, we still have reasons to believe that all the above studies represent efforts to explain the semantic function inherent in NP-*of* phrases by revealing how *of*-postmodifiers semantically relate to the head nouns. This interpretation, however, cannot evaluate the complexity on the phrasal level or explain how this phrasal complexity can work to promote the information flow over larger stretches of text, such as sentences and discourse. Biber, Gray, and Poonpon once tried to measure the semantic complexity of nominal phrases. They distinguished between prepositional postmodifiers with concrete/locative meanings and those with abstract meanings, associating these two types of postmodifiers with different degrees of complexity by placing the latter in a more advanced stage than the former (30–31). However, Ansarifar, Shahriari, and Pishghadam (2018) did not find any significant differences in the frequency of "of phrase (concrete/locative meanings)" and "of phrase (abstract meanings)" across three groups of abstracts written respectively by L1 Persian MA-level students, a group of L1-Persian doctoral students and a group of published writers within the field of Applied Linguistics (67–68). In another word, Biber, Gray, and Poonpon's method to assess semantic complexity of NP phrases may not be adequate enough to reflect learners' developmental progression and problems when they try to master the complexity features of NP-*of* phrases.

The other research direction involving NP-*of* phrases is the study of nominal lexical bundles. A wide range of lexical bundle studies focus primarily on comparing expert and non-expert writing to see how they differ in the function and structure

of lexical bundles (e.g., Hyland, 2009; Biber, Conrad, & Cortes, 2004; Biber & Barbieri, 2007; Chen & Baker, 2010; Pan, 2016). Lexical bundles of NP-*of* phrases are more frequently found in expert writing than non-expert writing, which has also led researchers to realize the noun-dominant feature of academic writing. However, the functions of nominal phrases, including NP-*of* phrases, are underexplored in the following two aspects:

1. The cut-off frequency requirement of lexical bundles excludes many useful but not that frequent nominal phrases, especially those with subject-specific content.
2. Lexical bundles of nominal phrases are mostly assigned the research-based or text-based function (see Hyland, 2009) without attention to their role in building argument over the discourse.

EAP instruction on how to use nominal phrases informed by the result of this frequency-based approach might lead to such a danger: EAP learners might be obsessed with top nominal lexical bundles and use the same grammatical structures profusely, but possibly for a different function from that which experts might use, to create the discourse (Cortes 142); for nominal phrases unqualified as lexical bundles in frequency, EAP learners will probably have little chance to realize their (lexical bundles') value to construct a compact and powerful academic discourse.

As can be seen, the above lexical bundle studies about the NP-*of* phrase mostly aim to explore how linguistic complexity is achieved with the high frequency of this structure, but generally overlook how this linguistic form of complexity operates semantically at different levels to meet the generic complexity of academic discourse. As Ryshina-Pankova has summarized based on Pallotti (2009)'s study, the problem is "complexity is studied as an ability to produce certain types of language forms, divorced of considerations of meaningful content production and realization of communicative goals" (52).

Thus, despite the vast scholarship on the NP-*of* phrase, there remains a need to systematically examine the discourse-semantic functions[1] this structure performs when it compresses information for complexity. Our study is also an effort to respond to the suggestion of Biber and Gray concerning the nominal complexity of academic discourse: "In order to succeed in an academic profession, students must learn to efficiently read and to eventually write this style of discourse. Future research is required to determine how that process can best be facilitated through EAP instruction" (19). To this end, the present study suggests a meaning-based method to measure the discourse-semantic functions of NP-*of* phrases in academic writing, and thereby conducts a comparative study between novice Chinese academic writers and more advanced academic writers to identify the former's possible problems in the context of thesis abstracts. Our functional and comparative analysis in the following part shows that EAP novice learners' use of NP-*of* phrases

---

[1]The term of "discourse-semantic function" is used interchangeably with "discourse-semantic complexity" in this study, for both are about the discourse-semantic aspect of linguistic complexity.

Table 1 Constituents of the two corpora

Corpus	Word count	No. of texts	Types	Type/token ratio
CNKI	34,974	100	3833	11.05
ProQuest	31,362	100	4537	14.55

in abstracts does not fit the communicative needs of their writing task as closely as more advanced learners. The aim of this study is to shed light on how EAP instructors can integrate the form and the meaning dimensions of NP-*of* phrases in order to teach the nominal academic style most productively.

3 Data

The investigative procedure takes the form of a contrastive corpus-based enquiry. The analyzed data consists of 200 English master's thesis abstracts, of which 100 were downloaded from CNKI, written by graduates from BUPT (they will be labeled henceforth as "CNKI corpus[2]"), 100 downloaded from ProQuest written by graduates mostly from universities in the USA, (they will be labeled henceforth as "ProQuest corpus"). The CNKI writers in this study are Chinese-speaking senior postgraduates majoring in computer science. They have all passed College English Test Band 4 (CET-4) and finished one general English writing course taught by Chinese instructors for 18 h in their first school year as postgraduates. The ProQuest writers in this study are all enrolled Computer Science postgraduates in universities of English-speaking countries, mostly the USA. As we mentioned before, ProQuest writers may have received more academic training and professional assistance from their school during their preparation of the theses. And the fact that the Dissertation and Thesis collection of ProQuest has been declared the official US off-site repository of the Library of Congress further led us to believe that the selected ProQuest writers of English can represent a higher level of academic writing proficiency than their CNKI counterparts. And to ensure we had access to the latest learners' writing materials, we started with CNKI looking for abstracts with distinct topics during the interval between 2010 and 2015. Then to maximize the comparability, we paired each abstract in CNKI with one having similar key words in ProQuest, since it is rare to find two abstracts with similar or identical topics in two different sets of data. The time interval between two abstracts in one pair was set to no more than 10 years to minimize any impact time might have on the comparison. The size of each finalized corpus for investigation is approximately 30,000 words, though the CNKI corpus is slightly longer than the ProQuest corpus in average length of abstracts and the latter is more varied in word choice (see Table 1).

---

[2] Sometimes we might use "CNKI abstracts"/"CNKI writers" and "ProQuest abstracts"/"ProQuest writers".

After retrieving all the concordance lines containing NP-*of* phrases in the two corpora using the corpus tool WordSmith 7.0 (Scott, 2007), we manually excluded the following NP-*of* phrases: (1) those for numbers and amounts, such as *a number of*, for the *of* part in these phrases has become a head noun rather than a postmodifier; (2) specialized terms, such as *Internet of Things* and *multicore scheduling of application threads*, for their occurrence has little to do with the writer's deliberate choice. Additionally, each occurrence of NP-*of* phrases was manually coded to differentiate between prepositional phrases functioning as noun postmodifiers versus all other functions (e.g., adverbials). The numbers of NP-*of* phrases before and after data refinement are found and shown in Table 2 below.

As Table 2 shows, CNKI writers, contrary to expectations, surpass their American counterparts in the frequency of NP-*of* phrases per 1000 words after refinement, which justifies the need for a qualitative analysis of NP-*of* phrases in the following sections.

4   Methodology

Our qualitative analysis is mainly built on Ryshina-Pankova's proposal of "using the concept of grammatical metaphor (GM), typically realized as nominalizations, as a construct that links linguistic complexity to discourse-semantic complexity and helps capture complex meaning making in advanced L2 writer text" (52). Ryshina-Pankova attributes the discourse-semantic complexity of GMs as nominalizations to its "conceptual refiguration of experience" and "configuration or development of concepts" (57). "Conceptual refiguration of experience" can help us understand the complexity of nominalization itself as a function that enables a generalization from individual cases to "abstract categories" to be evaluated and understood beyond the specific text. Since the quantity of nominalization alone is not adequate to measure the quality of the essays (Ryshina-Pankova 59), Ryshina-Pankova goes beyond the phrasal level, using "configuration or development of concepts" to explain how nominalization meets the complexity demand of the sentence and discourse. This high level function can "allow additional semantic density to cluster around them (nominalizations)" (55) and "configure these concepts into a particular line of reasoning in explanatory and argumentative texts" (58). This theorization of complexity views nominalization as one means to realize semantic complexity with the function of abstraction at the phrasal level and to realize discoursal/contextual complexity with the function of strategically placing nominalization in the sentence or

Table 2 Number of NP-of phrases in the two corpora before and after refinement

Before refinement		After refinement	
Corpus	No. of NP-*of* (tokens)	No. of NP-*of* (tokens)	Frequency per 1000 words
CNKI	1189	864	24.7
ProQuest	925	543	17.3

the text for the communicative appropriateness of the writing task. This approach to complexity offers us a tool to evaluate the effectiveness of NP-*of* phrases in producing abstract/condensed and argumentative meaning needed by the genre of academic abstracts. Thus, based on Ryshina-Pankova's theory, we propose the following meaning-based approach to measure the discourse-semantic functions of NP-*of* phrases in the two corpora.

Regarding the ways that the semantic complexity of NP-*of* phrases can function at the phrasal level, we try to measure it in terms of whether a nominalized relationship between the head noun and *of*-structure exists. We start by classifying the semantic relationship between the head noun and its *of*-phrase modifiers with a goal to see how the two corpora differ in abstracting and generalizing about concrete entities. For fear that frequency alone cannot display the full picture of novice academic writers' difficulties, we also look into the top ten head nouns in each corpus, hoping to capture any meaning-related difference in the context of abstracts.

For the function of discoursal complexity, we follow M. Ryshina-Pankova's practice to assess the function of nominalization with "the creation of links between the concepts" (54) in context-related terms but employ Halliday and Martin (1993)'s framework for evaluation. Both approaches can help judge the role of nominal phrases in making semantic connections and building argument in academic discourse. However, Ryshina-Pankova (2015) demonstrates his model using a case study without providing a concrete tool to assess a set of data with different topics while Halliday and Martin (1993) offer a more tangible and applicable sentence pattern as the criterion which better suits a corpus-based study dealing with different writing themes, especially in scientific discourse. Halliday and Martin argue that "the nature of scientific discourse which had to proceed step by step, with a constant movement from 'what we have established so for' to 'what follows from it next'; and each of these two parts, both the 'taken for granted' part and the new information, had to be presented in a way that would make its status in the argument clear" (81). They believe that for this purpose, the best way is to build the whole process into a single sentence with two nouns at each end and a verb setting up different "logic-semantic relations" between the two (Halliday and Martin 90). "This pattern obviously provides a powerful resource for constructing and developing an argument…construes iconically the total reality in which we now live, a reality consisting of semiotic entities in a periodic flow of information—a flow that one might well say has now become a flood" (Halliday and Martin 92). Therefore, we have reason to argue that NP-*of* phrases as either the subject or the object of this pattern may contribute more significantly to creating a powerful argument and a smooth information flow than those NP-*of* phrases in other patterns, and thus would consider this use of the NP-*of* phrase in a sentence or discourse as important to realizing its discoursal complexity. It is worthwhile to explore how this meaning-making function of NP-*of* phrases differs between academic writers at different levels.

So far, informed by Ryshina-Pankova's and Halliday and Martin's work, we have categorized different discoursal-semantic functions of NP-*of* phrases, and proposed a method of measurement. The following comparative analysis proceeds to find whether NP-*of* phrases in two corpora differ in discoursal-semantic functions

and how the differences might impact the realization of communicative needs of abstracts.

5 Comparison of Semantic Function Within NP-*of* Phrases Between the Two Corpora

In order to capture how nominalization functions semantically within the phrasal level, we firstly need to classify the internal meaning relationships within NP-*of* phrases. Drawing upon the previous classification of Quirk, Sidney, and Geoffrey (1275–1285), we listed five types of semantic relations between the head noun and its *of*-phrase which can cover nearly all NP-*of* phrases in the two corpora: (1) part-whole; (2) verb-object.; (3) subject-verb; (4) apposition; (5) noun-qualitative attributes. Examples are as follows:

1. Part-whole: e.g., the roof of this house (Quirk et al. 1277)
2. Verb-object: e.g., the imprisonment of the murderer (Quirk et al. 1278)
3. Subject-verb: e.g., the arrival of the train (Quirk et al. 1278)
4. Apposition: e.g., the news of the team's victory (Quirk et al. 1284)
5. Noun-qualitative attributes: e.g., a picture of a house (Sinclair 129)[3]

A chi-square test indicates that the two corpora differ significantly in terms of the distribution of these five types of NP-*of* phrases at the 0.05 level (*p*-value <0.00001) which is shown in Table 3. All cells, except for the cell of subject–verb relationship, have an absolute R value lower than 1.96, which suggests that the other four types of NP-*of* phrases contribute to the significant differences between the two corpora. The *R* values of the other four types can lead us to suggest that CNKI writers use significantly more NP-*of* structures with part-whole relationships and noun-modifier relationships, and fewer NP-*of* structures with verb–object relationships as well as appositive relationships than ProQuest writers.

The two types of NP-*of* structures CNKI writers use more frequently, part-whole and noun-qualitative attributes, happen to be relatively noun-oriented without obviously involving a unified process, which can be demonstrated in the following examples of CNKI corpus[4]:

1. In the design and implement phrase, this paper firstly introduced *the main function* of each module. (CNKI-28)

---

[3] Quirk only gave examples of noun-quantitative attributes, though he mentioned partitive constructions of both qualitative and quantitative attributes in his study (Quirk 1278). So we turned to Sinclair's study for examples of qualitative attributes.

[4] All the examples from the corpus will be identified by the corpus name followed by a number representing the location of the source abstract in the corpus, e.g., CNKI-01 means the example comes from no.1 abstract in CNKI Corpus. Both the number and the title of the source abstracts can be found in the Appendix.

Table 3 The contingency table for the distribution of semantic types of NP-of phrases within the phrasal level in ProQuest and CNKI

Corpus	Values ($p < 0.00001$)	Part-whole	Verb–object	Subject–verb	Apposition	Noun-qualitative attributes	Row totals
ProQuest	Observed count	198	197	46	57	45	543
	Expected count	232.33	152.44	45.54	35.51	77.19	
	R value	−2.3	3.6	−0.1	3.6	−3.7	
CNKI	Observed count	404	198	72	35	155	864
	Expected count	369.67	242.56	72.46	56.49	122.81	
	R value	1.8	−2.9	0.1	−2.9	2.9	
Column totals		602	395	118	92	200	1407 (grand total)

The results in Table 3 were generated by SPSS Statistics 22

2. The creativity of this paper about the *method of information security risk evaluation* lies in the following two factors. (CNKI-90)

   Even though some nominalized verbs, such as "evaluation" in (2) may occur in an *of*-structure of these two kinds of NP-*of* phrases, they merely provide ancillary semantic content to the head noun and do not semantically engage with it as directly and dynamically as those nominalized verbs with either their subject, object, or appositive.
   Conversely, the other two types of NP-*of* structures CNKI writers use less frequently, verb–object and apposition, mostly involve a nominalized process created by the head noun and the *of*-structure, as in the following examples from the ProQuest corpus:

3. The implementation of these devices on an enterprise can save on costs and add capabilities previously unavailable. (ProQuest-70)
4. …the needs of finding a friendly, comfortable and smart interface between users and *home environments* are growing. (Proquest-100)

   The marked difference in frequency of semantic types of NP-*of* phrases may show the CNKI corpus is generally less nominalized than the ProQuest corpus in the relationship between the head noun and its *of*-structures within NP-*of* phrases. The difference suggests that semantic complexity within NP-*of* phrases contained in the CNKI corpus is weaker than that in the ProQuest corpus, and thus may result in a less condensed style in CNKI abstracts.
   To further examine different semantic functions that NP-*of* phrases perform in the two corpora at the phrasal level in context-related terms, we also compared the top ten head nouns in NP-*of* phrases within each corpus. To avoid idiosyncrasies

Table 4 Frequency of top NP-of phrase head nouns in the two corpora

Number	CNKI top head nouns in NP-*of* phrases/frequency	ProQuest top head nouns in NP-*of* phrases/frequency
1	**Development/65**	Use/13
2	**Analysis/25**	**Problem/10**
3	**Characteristics/23**	Properties/10
4	Application/23	**Development/10**
5	**Performance/22**	**Analysis/9**
6	**Problem/16**	**Performance/8**
7	Process/15	**Characteristic/8**
8	**Method/14**	**Method/8**
9	Implementation/14	Quality/7
10	Basis/13	Results/7
Sum (percent of all the occurrences)	240 (240/864 = 27.8%)	93 (93/543 = 17.1%)

The bold-faced head nouns are those the two corpora have in common

from individual writers, we apply the criterion that the same head noun with its *of*-structure has to occur in at least 3–5 texts (a criterion to avoid writers' idiosyncrasies adopted in lexical bundle studies: e.g., Biber & Barbieri, 2007; Cortes, 2002; Chen & Baker, 2010) and found the CNKI corpus greatly exceeds the ProQuest corpus in the total frequency of the top head nouns as indicated in Table 4:

The statistics in Table 4 convinced us that compared with ProQuest writers, CNKI writers tend to overuse certain head nouns in NP-*of* structures most of which are nominalizations. In order to find the semantic function these overused NP-*of* phrases may serve, we try to locate them in abstract moves.

With a similar frequency criterion we have used to extract the top NP-*of* head nouns (occurrences in the same move of three abstracts), we found that the top NP-*of* head nouns in the CNKI corpus seem to be more closely associated with certain abstract moves than those in ProQuest, as the following table shows (five moves are represented by B for background, P for research problem, M for method, R for results, and C for conclusion):

The result in Table 5 shows that in the CNKI abstracts, the top NP-*of* head nouns semantically linked to the background and method move greatly outnumber those in the ProQuest abstracts (73/5; 40/0, respectively), which might be proof that certain NP-*of* phrases in the CNKI corpus are not used where abstraction is expected to be necessary but as "formulaic expressions" (Hyland 43) or "learner bundles" (Chen Yu-hua and Baker 31). This is confirmed by Hyland (2009) who has found that postgraduates tend to use more formulaic expressions than native speakers in order to prove their writing skills (Hyland 55).

Next, we try to specify the semantic content these formulaic NP-*of* phrase head nouns in the CNKI corpus tend to convey in order to pin down CNKI writers' problems.

Table 5 Top NP-of head nouns associated with each move in the two corpora

CNKI					ProQuest				
B	P	M	R	C	B	P	M	R	C
Development (63) Application (10)	Problem (9) Method (9) Implementation (11) Process (7)	Analysis (17) Characteristics (10) Basis (13)	Performance (12)		Quality (5)	Problem (8) Properties (6) Analysis (7)		Results (5)	
73	36	40	12	0	5	23	0	5	0

The number in the bracket represents the occurrences of the above top NP-*of* head noun in the corpus

In the background move, "development of" and "application of" are found to usually introduce generalized information, as illustrated in the following examples from the CNKI corpus:

5. With t*he rapid development of global economy and people's life rhythm speeding up, along with the rapid growth of the world aviation transport,* air transport is becoming increasingly busy, the demand of air monitoring control technology is rising, navigation system is facing new challenge. (Background move of CNKI-01)
6. With *the application of advanced science and technology*, the technical content of the various products is increasing, which puts forward higher requirements for maintenance training. (Background move of CNKI-29)

For the background move in the ProQuest abstracts, there is little generalized information about society, people's lives, or the whole industry but information about a related area of the industry, specifically on its contribution or needs. This difference may result in the more diverse choices of nominal phrases we find in ProQuest abstracts, as shown in the following examples from the ProQuest corpus:

7. As the tremendous growth of networks and e-business changes the nature of traditional information security threats, the capability to provide prompt and accurate information to authorized users boosts the competitive power of an organization. (Background move of ProQuest-42)
8. In this Big data era, the need for performing large-scale computation is evident. *A better understanding of the most suitable platforms* which can efficiently run these computations is needed. (Background move of ProQuest-87)

In the research problem move, there is no drastic difference in the frequency of top NP-*of* head nouns between the two corpora as Table 5 indicates, and the different choices of concrete words like "method" or "process" might be topic-related. However, we cannot help noticing the only nominalization in this move—"implementation"—which leads an NP-*of* phrase 11 times in the research problem move of the CNKI abstracts but zero time in the same move of the ProQuest abstracts. Instead of using "implementation," ProQuest writers are found to prefer the more congruent form-—"implement"—whenever the idea is needed in the research problem move of the of abstracts as the following examples show:

9. Also, this paper focuses on the implementation of clients function modules and finishes many tests by connecting with the server. (Research problem move of CNKI-41)
10. This thesis's objective is to develop and *implement* a low cost autonomous robot tank with Microsoft Kinect for XBOX as the primary sensor, a custom onboard small form factor (SFF), high performance computer (HPC) with NVidia GPU assistance for the primary computation, and OpenCV library for image processing functions. (Research problem move of ProQuest-32)

Regarding the method move, "basis of," "analysis of," and "characteristics of" common in CNKI are found mostly after prepositions like "on" and "according to," as illustrated in the following examples:

11. On the basis of getting the point of intersection between the two-dimensional free-form curves, research and analysis the envelope generate algorithm for producing new envelops. (Method move of CNKI-38)
12. Second, based on research and analysis of the popular research database connections for different development platforms, this paper proposed one abstract model and the overall design of the interface system, which included related seven main modules. (Method move of CNKI-42)
13. This dissertation designs a flexible privacy protection scheme based on application scenario and a query-based protection scheme *according to the characteristics of* Internet of Things. (Method move of CNKI-100)

These patterns are not found in the ProQuest abstracts which tend to offer a more direct and active way to present the method, such as:

14. *Based* on a plane-wave expansion approach, I proposed a method by which one can obtain the effective cladding indices by solving the full-vector wave equation for a periodic cladding structure. (Method move of ProQuest-93)
15. In this thesis, we propose a pub-sub network architecture design called MEDYM, for Match-Early with Dynamic Multicast. MEDYM follows the end-to-end distributed system design principle, by decoupling the content-based pub-sub service into a computationally intensive, application-specific event matching at network edge, and a simple, generic multicast routing in the network. (Method move of ProQuest-91)
16. To evaluate the feasibility of the mechanism with different types of AD sensors, we *analyze* its performance with content and volumetric sensors that characterize the payload and other relevant features of the traffic exchanged by a user. (Method move of ProQuest-85)

In the result move, "performance of" seems to be favored by CNKI writers when they try to prove the validity of their research, as the following examples demonstrate:

17. The experiments show that the solutions can effectively enhance the performance of the network. (Result move of CNKI-09)
18. Finally, to verify *the performance of* the proposed Lexicon-SVM method by micro-blog corpus, the results show Lexicon-SVM method shows higher accuracy. (Result move of CNKI-06)

As Table 5 shows, ProQuest writers prefer "results of" in the result move to reveal what they have found. Examples are:

19. *The results of this study* identify the different importance levels of security controls and sub-security controls. (Result move of ProQuest-90)

20. The results of this research include the release of the first open-source Android enterprise monitoring solution of its kind, a comprehensive guide of data sets available for collection without elevated privileges, and the introduction of a novel design strategy implementing various Android application components useful for monitoring on the Android platform. (Result move of ProQuest-70)

When ProQuest writers need to verify their findings, they often offer specific information about how the performance is improved without mentioning the improved performance at all, as in the following example:

21. The results show that CCN-FOH provides for a better user experience while keeping its overheads comparable to, if not better than, the less costly CCN-Publication. (Result move of ProQuest-99)

The above semantic analysis of the formulaic NP-*of* head nouns in the CNKI abstracts lead us to believe that they often introduce NP-*of* phrases which do not produce "useful complexity" (Ryshina-Pankova 60). In another word, the use of these NP-*of* phrases is not motivated by an expression of semantic complexity entailed by the writing task but an attempt of novice academic writers to exhibit some formal rules they believe apply to the writing of academic abstracts. Part of these rules in the case of novice writers in BUPT might derive from some so-called "universal" model or sample abstracts which tend to associate specific moves with certain expressions, such as some NP-*of* phrases. As the examples demonstrate, these move-specific NP-*of* phrases usually produce generalized information devoid of any specialized content. Instead of facilitating the information flow in abstracts, these nominal phrases only slow readers down by distracting them with unrelated information.

In a nutshell, the lower frequency with which nominalization occurs within NP-*of* phrases as well as the bigger number of formulaic expressions—move-specific NP-*of* phrase head nouns—might prevent CNKI writers from using a wider range of NP-*of* phrases to compress relevant information efficiently and flexibly within the phrase level. The consequence is that despite the dense use of NP-*of* phrases, the abstracts by novice academic writers may be still less compact than their academic communities expect. The next question is how NP-*of* phrases semantically relate to other parts of the sentence in order to function effectively in discourse construction.

6   Comparison of Discoursal Complexity of NP-*of* Phrases Between the Two Corpora at the Clausal Level

It is clear that in the dominant pattern in scientific argument proposed by Halliday and Martin (1993), verbs rather than conjunctions or prepositions are used to realize logical-semantic relationships between two processes represented by nominal phrases. Though they believe that the relationship within the dominant pattern is

Table 6 Common types of logical-semantic relation, with typical realizations as conjunction and preposition

Expansion type	Category	Typical conjunction	Typical preposition
Elaborating	Expository Exemplificatory	In other words I.e. For example; e.g.	Namely Such as
Extending	Additive Alternative Adversative	And Or But	Besides Instead of Despite(in contrast)
Enhancing	Temporal Causal Conditional Concessive Comparative	Then (at that time) So (for that reason) Then (in that case) Yet So (in that way)	After Because of In the event of Despite (contrary to expectation) Like

typically a cause-proof realized by verbs, such as "depends on," "accelerates," and "produce," Halliday and Martin still think it is an oversimplification, so they propose a model to realize the common types of logical-semantic relation by using conjunctions, prepositions and verbs as in Tables 6 and 7 (90–91):

Table 6 shows semantic relationships realized with prepositions or conjunctions. Table 7 shows the same relationships realized by a verb linking two nominalizations. We plan to compare between the two corpora the proportion of NP-*of* phrases either in the subject or object positions in the patterns of Table 7 in an effort to find any difference in the discoursal functions NP-*of* phrases serve at the clausal level. All these NP-*of* phrases in the subject or object positions in the lexicalized relationships are categorized and marked in Table 7. There are also NP-*of* phrases in the two corpora which do not interact directly with another noun-depicted process in a relationship lexicalized by a verb. They are:

1. NP-*of* phrases as part of adverbials, such as "with the development of the industry";
2. NP-*of* phrases which occur as objects in the infinitive or participle phrases, such as "to improve the efficiency of," "…thus ensuring the independence of the site data";
3. NP-*of* phrases which serve as subjects of intransitive verbs, such as "the application of GIS have made great progress" or the subject in a passive-voice clause with no agent given, such as "the complexity of space and time is reduced";
4. NP-*of* phrases which function as objects of verbs initiated by nouns that cannot depict processes, such as "we," "the paper," and "the study".

These categories of NP-*of* phrases are marked as 1, 2, 3, 4, respectively in Table 8.

The above categorization of NP-*of* phrases can help us find how they interact with other parts of the sentence to which they belong and know what role they play in connecting concepts in academic discourse. Since there are fewer than five counts in type 54, 55, 56, 58, 59, 510 in two corpora, we only calculate the occurrences of the other types in the following table:

Table 7 Examples of Lexicalization of logical-semantic relations (as verbs)

Category	Examples of lexicalization (verbs)
Expository (51)	Be represent constitute comprise signal reflect
Exemplificatory (52)	Be exemplify illustrate
Additive (53)	Accompany complement combine with
Alternative (54)	Replace alternate with supplant
Adversative (55)	Contrast with distinguish
Temporal (56)	Follow precede anticipate co-occur with
Causal (57)	Cause produce arise from lead to result in prove
Conditional (58)	Correlate with be associated with apply to
Concessive (59)	Contradict conflict with preclude
Comparative (510)	Resemble compare with approximate to simulate

Table 8 Comparison of frequency of NP-of phrases in different semantic relationships at the clausal level between the two corpora

Types of NP-*of* phrases	CNKI	ProQuest	Fisher's exact test
1	273	110	Yes ($p = 0.0004$)
2	74	38	No ($p = 0.36$)
3	26	10	No ($p = 0.23$)
4	232	104	Yes ($p = 0.0099$)
51	118	100	Yes ($p = 0.044$)
52	25	10	No ($p = 0.29$)
53	6	15	Yes ($p = 0.003$)
57	125	152	Yes ($p = 0.00$)
Sum	859	537	

Results show that compared with the ProQuest abstracts, the CNKI abstracts use significantly more NP-*of* phrases as part of adverbials in type 1 and more NP-*of* phrases as objects of sentences with subjects like "the paper," "we," and "the study" in type 51; and much fewer NP-*of* phrases in a cause–effect relationship in type 57. In order to more closely fit Halliday & Martin's model that highlights nominalization, our investigation is narrowed down to NP-*of* phrases which involve nominalizations. Among all the NP-*of* phrases in the verbalized cause–effect relationship, 92 out of 152 (61%) cases in ProQuest involve nominalizations compared to 49 out 125 (32%) in CNKI. That means the difference between two corpora grows even bigger when we focus on the frequency of nominalized NP-*of* phrases occurring in a cause–effect relationship.

Halliday and Martin discovered that "scientific writers, from Issac Newton onwards, increasingly favour such a mode of expression—one in which, instead of writing 'this happened, so that happened', they write 'this event caused that event'" (81). Thus, compared with ProQuest writers, CNKI writers are less able to build knowledge in a scientists' way in terms of using NP-*of* phrases at the clausal level. For ProQuest writers, the prevalence of NP-*of* phrases, especially nominalized ones in a cause–effect relationship in their abstracts, shows that more advanced writers

are better at creating more discoursal complexity with *NP-of* phrases by connecting two processes within one clause to efficiently push forward the discussion.

Carroll further argues that a congruent expression of cause–effect means "there is very little argumentation here" since "because gives no sense of the type of role the cause played" (16). In this sense, NP-*of* phrases in the ProQuest abstracts seem to create more argument than those in the CNKI abstracts at the clausal level.

The next step is to find the moves that contain these nominalized NP-*of* phrases in a causal-effect relationship (shortened as nominalized causal NP-*of* phrases hereafter), which will shed light on any discoursal effects they might have on the communicative demands abstracts exert.

7   Comparison of Discoursal Complexity of NP-*of* Phrases Between the Two Corpora at the Discoursal Level

Numbers in Table 9 reveal that both the CNKI and ProQuest abstracts mainly use nominalized causal NP-*of* phrases in the background move but a major difference emerges in the problemizing step of this move. Of nominalized causal NP-*of* phrases in the background move, only 7 out of 28 (25%) in CNKI occur in the problemizing step while in ProQuest 20 of 43 (46.5%) occurrences happen in the same step. Another difference between the two corpora is that CNKI writers use nominalized causal NP-*of* phrases in the result move and conclusion move considerably less often than ProQuest writers (8.2% vs. 25%; 8.2% vs. 13%, respectively).

Swales and Feak noted that "Academics are famous for problemizing….We often do this in order to justify or prepare the way for our own research" (23). Compared with the generalized background information, the problemizing step in the background can more effectively highlight the value of the research by revealing a gap and leading naturally to the relevance of the research question. In this sense, a cause–effect argument using NP-*of* nominalizations in this step can efficiently and straightforwardly accentuate the connection between the social needs and the research problem in an argumentative manner as the following examples illustrate:

22. However, *estimation of cyclic features under constrained sensing time* suffers from *cyclic frequency offsets resulting from non-synchronous sampling and local oscillator offsets*. (Problemizing step of ProQuest-65)
23. …the *detection of application-specific vulnerabilities* necessitates *the examination of both logic flows among web pages and user inputs of each page*. (Problemizing step of ProQuest-100)

Table 9 Distribution of nominalized clausal NP-of phrases in different moves of the two corpora

Corpus	B	P	M	R	C	Sum
CNKI	28	0	13	4	4	49
ProQuest	43	4	8	23	12	92

Apart from the problemizing step in the background move, the result and conclusion of abstracts are also key moves in that they address the questions of "what was discovered" and "what do the findings mean?" (Swales & Feak, 2009, p. 5). In these two parts of abstracts which are essential to sell the research, the cause–effect relationship realized with the help of NP-*of* nominalizations can create a stronger and clearer argument and thus, can help convince readers of the validity and value of the research, as is shown in the following examples:

24. The reduction in complexity and energy cost comes from *discretization of feature* space based on tolerable frequency offsets for the required classification accuracy. (Result move of ProQuest-65)
25. The *combination of all the innovations* allows building a robust and fully automatic human face processing system. (Conclusion move of ProQuest-91)

The increased argument in the problemizing step, the result and the conclusion of abstracts is also consistent with Hyland's data which shows that "major increases in the percentage of abstract moves between 1980 and 1997 occurred in the opening and closing moves" (Swales & Feak 2009, p. 20). Hyland believes that these changes are due to a stronger need to attract readers in response to the increased competition to have papers accepted by journals (Swales & Feak 2009, p. 20). Therefore, we have reason to believe that the more frequent use of NP-*of* phrases in the cause–effect pattern in the ProQuest abstracts may turn out to be one means that more advanced academic writers employ to lend support to self-promotion in the genre of thesis abstracts. Novice academic writers represented by CNKI writers in this study, however, seem less able to employ NP-*of* phrases vigorously for this function.

8   Conclusion

From a meaning-based point of view, this study adopts a combination of quantitative and qualitative approaches in an attempt to assess the discourse-semantic complexity of NP-*of* structures and detect how it differs between academic writers at different levels of proficiency. Taking English-language abstracts within the same discipline of computer science by two different groups of learners as samples, we find that with regard to semantic complexity, novice writers represented by the CNKI corpus show less flexibility in realizing this function by underusing NP-*of* phrases with a nominalized relationship and, at the same time, overusing some head nouns in NP-*of* phrases which they might mistakenly associate with certain moves. Conversely, ProQuest writers tend to use more nominalized NP-*of* phrases and individualize their word choices of NP-*of* head nouns to more flexibly categorize concrete experiences into abstractions. Regarding the discoursal complexity, CNKI writers tend to use more NP-*of* phrases as part of temporal adverbials or objects governed by static subjects like "the paper," "we," or "the study" to report the content of their master theses. By contrast, ProQuest writers use far more NP-*of* phrases in a cause–effect relationship to reason with readers by linking different

processes, especially in the problemizing step, result move and conclusion move, where arguments are most needed for self-promotion. Our finding suggests that though CNKI writers use the NP-*of* phrase even more frequently than ProQuest writers, they may have the tendency to treat it as a mere formal feature of academic writing, and thus fail to relating its use to the communicative requirement of their writing task as closely as ProQuest writers. Specifically speaking, CNKI writers are firstly found to be less able to employ the NP-*of* phrase diversely to nominalize discipline-specific processes as expert readers in their academic community may anticipate in abstracts, and secondly they are weaker in forging cause–effect reasoning to improve the persuasiveness of their research by using nominalized causal NP-*of* phrases in important abstract moves.

These problems, however, should not be viewed as errors but need to be treasured as the developmental features in their process of mastering the noun-dominated feature in academic writing. EAP teachers are suggested to take the discourse-semantic view of linguistic complexity into consideration. They not only need to lead novice academic writers to realize the heavy presence of NP-*of* phrases in academic writing but also have to guide them into using this structure to respond to the complexity of their academic writing tasks. In the actual class, teachers may provide extensive practice in decoding discourse-semantic functions of NP-*of* phrases in representative abstracts relating to students' fields, encourage students to collect discipline-specific processes usually nominalized into NP-*of* phrases, and try to identify the cause if students overuse certain NP-*of* phrases in some abstract moves.

Further research is needed to reveal the progress that novice academic writers exhibit when they struggle to grasp the functions of NP-*of* phrases in actual pedagogical practice. Last but not the least, the major limitation of this study is that both corpora represent learner language, which cannot guarantee a clear distinction between two writing levels. It would be a productive area for future study to compare the discourse-semantic complexity of NP-*of* phrases between student writing and the expert writing.

Appendix: Composition of the Corpora in This Study

Number of the source abstract	Title
CNKI-01	Communications and navigation monitoring coverage simulation system by using GIS
CNKI-02	Emergency dispatch system based on GIS design and implementation
CNKI-03	A software implementation of electromagnetic simulation and automated testing platform
CNKI-04	Efficient transmission for multi-type information in space tracking network
CNKI-05	Design and implementation of UE-based recommendation system

Number of the source abstract	Title
CNKI-06	The design and implementation of log analysis system for the user-oriented personalization recommendation
CNKI-07	The design and implementation of information extraction and storage system on video vertical search engine
CNKI-08	Research and implementation of dynamic face recognition technology based on mobile video
CNKI-09	Research and design of video information electromagnetic leakage suppressor
CNKI-10	Program design and application of the value-added business operations platform
CNKI-11	The design and research of a RFID reader antenna
CNKI-12	Research on LTE-A system of the base station coordination scheduling
CNKI-13	Research based on key technologies in vehicular AD-HOC network
CNKI-14	A research on VANET data transmission mechanism based on network topology dynamics analysis
CNKI-15	Development of key technologies in fingerprint-based mobile positioning
CNKI-16	EEG applied research in personal identification and fatigue detection
CNKI-17	Design and implementation of automated testing tools for Android application
CNKI-18	Development of location information services in Android platform based on Baidu Maps
CNKI-19	Study on privacy preserving data mining under horizontal distribution
CNKI-20	Research and implementation on recommendation algorithm based on location and trust of food businesses
CNKI-21	Research on joint resource allocation in LTE-advanced based Internet of vehicles network and building of application platform
CNKI-22	A research and implementation of public transportation path optimization algorithm based on traffic condition
CNKI-23	The research and development on vehicle and crew scheduling problem based on intelligent optimization algorithms
CNKI-24	Mobile intelligent recommendation platform for public transportation users
CNKI-25	Information and communication technology research based on cooperative vehicle infrastructure system
CNKI-26	The research and application based on OSGI Technology
CNKI-27	Network optimization based on data mining
CNKI-28	Design and development of social platform for Webserver
CNKI-29	The cognitive load research of human–computer interaction in the virtual maintenance training
CNKI-30	The design and implementation of data center system using for bottles anti-counterfeiting and traceability
CNKI-31	Research on CNC lathe tool parameters measurement based on image processing
CNKI-32	A software development kit for display and control of indoor maps
CNKI-33	The mobile map Mashup application system design and research based on Arcgis Server

Number of the source abstract	Title
CNKI-34	The research of IoT information system based on Webgis
CNKI-35	Design and application of spatial database based on Hypertable
CNKI-36	The design and implementation of OpenAPI for indoor map information service
CNKI-37	The design and implementation of indoor map construction system based on Arcgis engine
CNKI-38	Research and implementation of regional synthesis and envelope generation algorithm based on GIS platform
CNKI-39	The measuring of the engine vane molding surface based on grating projected method
CNKI-40	Detecting the overlapping community structure in social networks
CNKI-41	Design and development of intelligent spectrum sensing application based on Android platform
CNKI-42	Research and implementation of database unified interface system for wisdom oilfield applications
CNKI-43	New technology of image processing based on statistical analysis and its application in Photodynam
CNKI-44	Application layer virtual method based on software-defined network
CNKI-45	Design and implementation of consumer software for contactless IC card
CNKI-46	Some key technologies research in wireless communication system based on MIMO
CNKI-47	The implementation of Linux Kernel based on S+Core CPU
CNKI-48	Design of embedded control system for the riderless bicycle
CNKI-49	Research on global planning algorithm of mobile robot in complex and dynamic environment
CNKI-50	Research on security technology of Internet of Things
CNKI-51	Analysis and design of topic crawler
CNKI-52	Spam filtering based on online ranking logistic regression
CNKI-53	Research and implementation of load balancing technology for SDN network controller
CNKI-54	Research on time-domain interference coordination and resource allocation in heterogeneous networks
CNKI-55	Convolutional neural networks for hyperspectral image classification
CNKI-56	System identification and simulation of bicycle robot
CNKI-57	Face recognition platform design and the complementation of face recognition algorithm
CNKI-58	Design and implementation of dynamic virtual machine security domain control service
CNKI-59	Applications of SVM in data mining
CNKI-60	Research on information hiding algorithm based on the digital image
CNKI-61	Research and practice of video image quality test technology
CNKI-62	Data mining engine based on Big data
CNKI-63	The research of the influence on Chinese user acceptance behavior TOO20 E-Commence model

Number of the source abstract	Title
CNKI-64	Research and implement on data mining algorithm parallel based on Hadoop
CNKI-65	Generation and application of high quality optical frequency comb
CNKI-66	Design and implementation of multisource download system based on network crawling technology
CNKI-67	Capacity optimization for full-duplex massive multiuser MIMO systems based on random matrix methods
CNKI-68	Design and implementation of shopping online system
CNKI-69	Research on real-time global illumination mode in complex scenes
CNKI-70	Design and implementation of multi-location sharing application system based on Android mobile terminal
CNKI-71	The study of the security and privacy protection of Big data based on cloud computing
CNKI-72	Prediction of ads' click through rate based on recurrent neural network
CNKI-73	Case assistant system based on the intelligent video analysis technology
CNKI-74	Research on dual-band antenna over EBG Structure and band-notched UWB antenna
CNKI-75	Comparison of data mining techniques used to predict student retention.
CNKI-76	Study on key technologies of wireless optical communication
CNKI-77	Research on the construction of the Quasi-Cyclic low-density
CNKI-78	Analysis and design of high-level programming language assisted learning system
CNKI-79	Research on a new path finding algorithm of mobile robot in the complex environment
CNKI-80	The research of feature selection method and sentiment analysis based on microblog
CNKI-81	Radio resource management in broadband satellite communication systems
CNKI-82	Design and implement of taxi call system based on IOS
CNKI-83	Application of machine learning algorithms in data mining
CNKI-84	The study of models and theory researches of optimization of demand and supply in smart grid
CNKI-85	The research of constraint QoS routing technology based on the next generation network
CNKI-86	Research on recognition technology based on stationary cycle signal modulation
CNKI-87	The research and implement of data mining algorithms based on Hadoop
CNKI-88	Design and development of social platform for webserver
CNKI-89	Research of risk detection for Android application
CNKI-90	The research of information security risk assessment model and method
CNKI-91	Empirical study on effect model of podcasting core information releasing behavior
CNKI-92	The improvement and application of k-means algorithm
CNKI-93	Research on infrared absorption characteristics of photonic crystal and its application

Number of the source abstract	Title
CNKI-94	Coalition game-based cooperative spectrum sensing research in cognitive radio networks
CNKI-95	Design and optimization of camera application based on Android system
CNKI-96	Quantum state preparation and application in cavity QED
CNKI-97	Simulated annealing-based congestion control strategy
CNKI-98	The controller placement problem and cloud storage assignment problem in SDN
CNKI-99	The research of routing mechanism for named data network
CNKI-100	Towards and autonomous mobile control unit of smart homes
ProQuest-01	Improvement of security in UAS communication and navigation using ADS-B
ProQuest-02	A flexible simulation framework for processor scheduling algorithms in multicore systems
ProQuest-03	Scalable ray tracing with multiple GPGPUs
ProQuest-04	Compressing genome resequencing data
ProQuest-05	A new class of techniques for Web personalization
ProQuest-06	Accelerating Mahout on heterogeneous clusters using HadoopCL
ProQuest-07	Towards next generation vertical search engines
ProQuest-08	Simultaneous multi-view face tracking and recognition in video using particle filtering
ProQuest-09	Securing unfamiliar system entry points against faulty user authentication via electromagnetics
ProQuest-10	Web services for application development in next generation telecommunications networks
ProQuest-11	Long range ultra-high frequency (UHF) radio frequency identification (RFID) antenna design
ProQuest-12	Health and activity monitoring system
ProQuest-13	Routing protocol evaluation and development of a fully functional simulation environment for vehicles
ProQuest-14	Protocole de routage intelligent pour les réseaux ad hoc de véhicules
ProQuest-15	Indoor positioning system using software defined radio (SDR) and support vector machine (SVM)
ProQuest-16	Creating and breaking it: Design of an ERP biometric protocol and development of a device for ERP
ProQuest-17	Replay debugger for multi-threaded Android applications
ProQuest-18	CAP—a context-aware privacy protection system for location-based services
ProQuest-19	Data transformation for privacy-preserving data mining
ProQuest-20	Sparsity, scalability, and distribution in recommender systems
ProQuest-21	A framework for reliable and efficient communications in vehicular networks
ProQuest-22	Routing protocol evaluation and development of a fully functional simulation environment
ProQuest-23	Genetics-based concurrent planning and scheduling for flexible and modular manufacturing systems

Number of the source abstract	Title
ProQuest-24	Human activity detection using smartphones and maps
ProQuest-25	Flexibility in dependable real-time communication
ProQuest-26	Reconfigurable software defined radio platform an OSGi approach
ProQuest-27	A Tabu search, augment-merge heuristic to solve the stochastic location arc routing problem
ProQuest-28	A web application based on the MVC architecture using the Spring Framework
ProQuest-29	Design and implementation of human–computer interaction based on remote control for DRC system
ProQuest-30	A framework for data delivery in integrated Internet of Things architectures
ProQuest-31	Development of an automated system to evaluate the surface condition of grinding wheels
ProQuest-32	Mobile robot tank with GPU assistance
ProQuest-33	Exploring San Francisco's treasures: Mashing up public art, social media, and volunteered geographic information
ProQuest-34	Time-reversal massive multipath effect and bandwidth heterogeneity
ProQuest-35	Coupling environmental models and geospatial data processing
ProQuest-36	A study of mobile internet usage and the implications for mobile search interfaces
ProQuest-37	Mapping native plants A mobile GIS application for sharing indigenous knowledge in Southern California
ProQuest-38	Social media integration for open source GIS applications
ProQuest-39	Advanced multidirectional UV lithography for three-dimensional (3-D) micro-nano structures
ProQuest-40	Discovering community structure by optimizing community quality metrics
ProQuest-41	A feasibility study of distributed spectrum sensing using mobile devices
ProQuest-42	Distributed semantic web data management in HBase and MySQL cluster
ProQuest-43	Target—a new photo-mechanics measuring method based on digital image processing
ProQuest-44	Policy-driven network defense for software-defined networks
ProQuest-45	Design and research of ID card automatic recognition system
ProQuest-46	Precoding over the multiuser multi-input multi-output broadcast channel
ProQuest-47	Linux kernel module for security enhancement
ProQuest-48	Integrated motion control of wheeled mobile robots
ProQuest-49	New insights into bio-inspired intelligence: Safety-aware navigation and mapping of an autonomous vehicle
ProQuest-50	The insecurity of things: How to manage the Internet of Things
ProQuest-51	Eliminating SQL injection and cross-site scripting with aspect oriented programming
ProQuest-52	Regression with categorical predictors and random effect logistic models
ProQuest-53	Improving load balancing mechanisms of software-defined networks using open flow
ProQuest-54	Cost-effective deployment of heterogeneous networks and the effect of pilot contamination
ProQuest-55	Hyperspectral image classification using deep neural networks

Number of the source abstract	Title
ProQuest-56	Integrating centralized and decentralized approaches for multi-robot coordination
ProQuest-57	Applications of convolutional neural networks to facial detection and recognition for augmented reality and wearable computing
ProQuest-58	Modeling and analysis of security standards for webservices and cloud computing
ProQuest-59	Autonomous navigation algorithms for indoor mobile robots
ProQuest-60	An efficient approach to categorizing association rules
ProQuest-61	Multi-view face processing in video imagery
ProQuest-62	Research on application of data mining in intrusion detection security audit
ProQuest-63	How do organizations prepare and clean Big data to achieve better data governance? A Delphi study
ProQuest-64	Acceptance of health services on mobile phones: A study of consumer perceptions
ProQuest-65	Enhancements to the data mining process
ProQuest-66	Stable optical frequency comb generation and applications in arbitrary waveform generation, signal processing, and optical data mining
ProQuest-67	An enhanced malicious web crawler detection and classification system
ProQuest-68	On duality of MIMO relays and performance limits of full-duplex MIMO radios
ProQuest-69	The influences of atmospheric cues on consumer behavioral intentions: An affordance perspective
ProQuest-70	A system study of high dynamic range imaging
ProQuest-71	Android forensics: Automated data collection and reporting from a mobile device
ProQuest-72	Applying hybrid cloud systems to solve challenges posed by the Big data problem
ProQuest-73	Intelligent systems for quality defect prediction in injection molding
ProQuest-74	Adaptive sparse representations for video anomaly detection
ProQuest-75	Bandwidth enhancement of monopole UWB antenna with new slots and EBG structures
ProQuest-76	The research and application of data mining technology in analysis for students' performance
ProQuest-77	Polarization effects in optical fiber communication and distributed vibration sensing systems
ProQuest-78	Designing structured low density parity check codes with large girth
ProQuest-79	Characterizing the usage of computer-mediated communications
ProQuest-80	Sentiment analysis of Twitter data
ProQuest-81	Study on the modeling, analysis, and compensation of on-board processing (OBP) multicarrier demultiplexer/demodulators (MCDD) for the MF-TDMA signal in the presence of degrading factors
ProQuest-82	The taxi: Friend or foe? Understanding planners' perceptions of the taxi industry
ProQuest-83	Machine learning with incomplete information
ProQuest-84	Analysis and characterization of wireless smart power meter

Number of the source abstract	Title
ProQuest-85	Behavior-based admission and access control for network security
ProQuest-86	Wideband cyclostationary spectrum sensing and modulation classification
ProQuest-87	Performance comparison of two data algorithms on Big data platforms
ProQuest-88	Web archive services framework for tighter integration between the past and present web
ProQuest-89	Anomaly detection system using system calls for Android smartphone system
ProQuest-90	An approach to information system security assessment
ProQuest-91	Architecture design for distributed content-based publish-subscribe systems
ProQuest-92	An iterative cluster correction enhancement for the bisecting K-means algorithm
ProQuest-93	Photonic crystal fibers: Characterization and supercontinuum generation
ProQuest-94	Power control and multi-target identification in cognitive wireless networks
ProQuest-95	Creating and documenting Android web applications
ProQuest-96	Studies in quantum control and quantum entanglement
ProQuest-97	Routing optimization in wireless AD HOC and wireless sensor networks
ProQuest-98	An approach for mobility support in software-defined network-based mobile cloud computing
ProQuest-99	CCN forwarding strategies
ProQuest-100	Internet of Things based energy aware smart home control system

References

Ansarifar, A., Shahriari, H., & Pishghadam, R. (2018). Phrasal complexity in academic writing: A comparison of abstracts written by graduate students and expert writers in applied linguistics. *Journal of English for Academic Purposes, 31*, 58–71. https://doi.org/10.1016/j.jeap.2017.12.008

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

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

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes, 9*(1), 2–20. https://doi.org/10.1016/j.jeap.2010.01.001

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. https://doi.org/10.5054/tq.2011.244483

Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511805226

Cai, J.-g. (2003). *English writing*. Shanghai: Fudan University Publishing House. http://www.docin.com/p-1643660778.html

Carroll, J. E. (2016). The whole point of the thing: How nominalisation might develop students' written causal arguments. *Teaching History, 162*, 16.

Chafe, W. L. (1982). Integration and Involvement in speaking, writing, and oral literature. In D. Tannen (Ed.), *Spoken and written language: Exploring orality and literacy* (pp. 35–53). Norwood, NJ: Ablex.

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

Cortes, V. (2002). Lexical bundles in freshman composition. *Using Corpora to Explore Linguistic Variation, 9*, 131–145. https://doi.org/10.1075/scl.9.09cor

Coffin, C. (2009). *Historical discourse: The language of time, cause and evaluation*. Bloomsbury Publishing.

Halliday, M. A. (1988). On the language of physical science. *Registers of written English: Situational Factors and Linguistic Features*, 162–178.

Halliday, M. A. K. (2004). *An introduction to functional grammar (Third Edition revised by Christian Matthiessen)*. London: Routledge. https://doi.org/10.4324/9780203783771

Halliday, M. A. K., & Martin, J. R. (Eds.). (1993). *Writing science: Literacy and discursive power*. Washington, DC: Falmer Press.

Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency: Definitions, measurement and research. Language Learning & Language Teaching (MS).

Hunston, S. (2002). *Corpora in applied linguistics*. Ernst Klett Sprachen.

Hyland, K. (2009). *Academic discourse: English in a global context*. In. London: Continuum.

Kroll, B. (1977). Ways communicators encode propositions in spoken and written English: A look at subordination and coordination. In *Discourse across time and space* (pp. 69–108). Los Angeles: University of South California.

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

O'Donnell, R. C. (1974). Syntactic differences between speech and writing. *American Speech, 49*(1/2), 102–110. https://doi.org/10.2307/3087922

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics, 30*(4), 590–601.

Parkinson, J., & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for academic purposes students. *Journal of English for Academic Purposes, 14*, 48–59.

Pan, F. (2016). A corpus-driven structure and functional analysis of lexical bundles in journal articles written by English natives and Chinese writers. *Foreign Languages and Their Teaching, 4*, 115–123.

Ping, L., & Yan, C. (2012). The analysis and application of abstract nominal phrases in titles of scientific journal papers. *Scientific Writing, 23*(2), 322–324. Retrieved form http://dx.chinadoi.cn/10.3969/j.issn.1001-7143.2012.02.043

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). In R. Quirk (Ed.), *A comprehensive grammar of the English language*. London: Longman.

Ravelli, L. J. (1996). Making language accessible: Successful text writing for museum visitors. *Linguistics and Education, 8*(4), 367–387.

Ryshina-Pankova, M. (2015). A meaning-based approach to the study of complexity in L2 writing: The case of grammatical metaphor. *Journal of Second Language Writing, 29*, 51–63. https://doi.org/10.1016/j.jslw.2015.06.005

Sinclair, J. M. H. (2005). *Collins Cobuild English grammar*. Glasgow: Harper Collins Publishers.

Scott, M. (2007). *Oxford wordsmith tools (version 7.0) [computer software]*. Oxford: Oxford University Press.

Swales, J. M., & Feak, C. B. (2009). *Abstracts and the writing of abstracts* (Michigan series in English for academic & professional purposes) (Vol. 1). Michigan: Michigan University Press. https://doi.org/10.3998/mpub.309332

Yoon, H.-J. (2017). Linguistic complexity in L2 writing revisited: Issues of topic, proficiency, and construct multidimensionality. *System, 66*, 130–141.

Jing Li earned her master's degree in Linguistics from Beijing Foreign Studies University in 2002. Since her graduation, she has been working as a lecturer with School of Humanities, Beijing University of Posts and Telecommunications (BUPT). Her area of research is ESP discourse analysis and pedagogy. She has been committing herself to improving the academic writing and workplace communication skills of polytechnic postgraduates in BUPT through research and teaching. So far, she has published four journal papers, worked as one key member of a ministry-level project and principal investigator (PI) of a school project.