ISSN: 2710-0170

A Comparative Study of Lexical Bundles in the Abstracts of Chinese Master's Theses and Research Articles by English Expert Writers

Zihan Wu

Northwest Normal University, College of Foreign Languages and Literature, Gansu, Lanzhou, 730070, China

Abstract

This study compares the use of lexical bundles in the abstracts of Chinese master's theses and those of research articles written by English expert writers. Two small corpora were constructed: a corpus of master's thesis abstracts and a corpus of expert research article abstracts. Based on Hyland's (2008b) functional classification framework, four-word lexical bundles were extracted and functionally annotated using AntConc. The findings show that: (1) the normalized frequency of lexical bundles in the expert corpus is significantly higher than that in the master's corpus, indicating a greater degree of formulaic density among expert writers; (2) in terms of functional distribution, text-oriented bundles account for the highest proportion in the expert corpus (45%), reflecting stronger discourse organizational awareness, while research-oriented bundles account for the highest proportion in the master's corpus (42%), reflecting a focus on research background and methods; (3) master's thesis abstracts exhibit certain weaknesses, including a preference for verb-based bundles and vague quantifiers, overuse of certainty verbs, avoidance of first-person pronouns, and formulaic use of bundle patterns. This study provides empirical evidence for the teaching of academic English writing.

Keywords

Lexical bundles; Abstract; Academic writing.

1. Introduction

In academic English writing, lexical bundles, as frequently recurring multi-word sequences, constitute an important linguistic resource for constructing academic discourse. Lexical bundles are not only basic units of language production but also play crucial roles in discourse organization, stance expression, and the description of research activities. A substantial body of research has shown that the appropriate and flexible use of lexical bundles directly affects the fluency, idiomaticity, and persuasiveness of academic writing[1,2]. The abstract serves as the "face" of a research paper, undertaking the important functions of conveying key research information, attracting readers, and facilitating academic exchange. The characteristics of lexical bundle use in abstracts often reflect the author's language proficiency, academic writing ability, and genre awareness. In recent years, with the rise of corpus linguistics, lexical bundle research has received extensive attention in the field of second language writing. Existing studies have mostly focused on differences in lexical bundle use across different language proficiency levels, disciplinary backgrounds, or registers. However, comparative studies specifically targeting the important academic genre of the abstract remain relatively limited, and systematic comparisons between the abstracts of Chinese master's theses in English and those of journal articles by native English experts are particularly scarce. Against this background, this study aims to compare the use of lexical bundles in the abstracts of Chinese

ISSN: 2710-0170

master's theses and those of English expert writers. By constructing two small corpora and applying Hyland's functional classification framework[3], it systematically analyzes the similarities and differences in the distribution patterns and functional use of lexical bundles between the two types of abstracts, in order to reveal the characteristics and shortcomings of Chinese master's students' lexical bundle use in academic abstract writing and to provide a reference for the teaching of academic English.

2. Research Design

2.1. Analytical Framework

This study adopts Hyland'sfunctional classification framework of lexical bundles as its theoretical basis[3]. This framework divides lexical bundles in academic discourse into three major categories: research-oriented bundles, which help writers structure their research activities and describe the real world, including subcategories such as location, procedure, quantification, description, and topic; text-oriented bundles, which are concerned with the organization of the text and the development of argumentation, including subcategories such as transition signals, resultative signals, structuring signals, and framing signals; and participant-oriented bundles, which focus on the writer or reader, expressing stance or guiding reader participation, including the subcategories of stance features and engagement features. Hyland's framework accommodates the cognitive, organizational, and social-interactive dimensions of academic discourse, possessing strong explanatory power and operability. Therefore, it serves as the core analytical tool for functional analysis of lexical bundles in this study.

2.2. Research Questions

Based on the above analytical framework, this study proposes three research questions:

1) What are the similarities and differences in the overall distribution patterns of lexical bundles between the abstracts of Chinese master's theses and those of research articles by English expert writers?

2) What are the characteristic differences in the use of different functional categories of lexical bundles between the two types of abstracts?

3) What are the main weaknesses in the use of lexical bundles in Chinese master's thesis abstracts?

2.3. Research Methods

In terms of research methods, this study first constructed two small corpora. The corpus of master's thesis abstracts contains 10 English abstracts of Chinese master's theses in English. These papers were selected from the graduation theses of the applied linguistics departments of ordinary universities in five regions of China (Northeast, Northwest, Southeast, Central and Southwest). The publication period of the theses is from 2023 to 2025. This corpus totals 38,044 characters. The corpus of expert research article abstracts contains 10 English abstracts of research articles published in SSCI Q1 journals, including *Language Learning, Applied Linguistics, and The Modern Language Journal*, also published between 2023 and 2025. This corpus totals 13,828 characters. The size difference between the two corpora reflects the genre characteristic that master's thesis abstracts are generally longer than research article abstracts. In subsequent analyses, all frequency data are converted to frequency per million words (pmw) to control for corpus size effects. Lexical bundle identification employs a combination of frequency-driven and manual verification methods. The specific steps are: using the N-gram function of AntConc to extract four-word sequences from the corpora, setting a minimum frequency threshold of at least 3 occurrences per bundle in each corpus to balance corpus size

ISSN: 2710-0170

differences and extraction quantity, manually removing incomplete sequences and proper names, and finally functionally annotating all extracted bundles according to Hyland's functional classification framework [3].

3. Results and Discussion

Regarding the overall distribution of lexical bundles, after extraction and filtering, the master's abstract corpus yielded 78 bundle types with a normalized frequency of 2,050 per million words; the expert abstract corpus yielded 64 bundle types with a normalized frequency of 4,628 per million words. The data show that although the absolute number of bundle types in the master's abstract corpus is slightly higher than in the expert corpus, the normalized frequency of lexical bundles in the expert corpus is significantly higher. This means that within a unit of text, expert writers use lexical bundles with greater density, exhibiting a higher degree of formulaic language use.

In terms of functional distribution, the two types of abstracts show different characteristics. Research-oriented bundles account for the highest proportion in the master's abstracts (42%), followed by text-oriented bundles (38%) and participant-oriented bundles (20%). In the expert abstracts, text-oriented bundles account for the highest proportion (45%), followed by research-oriented bundles (38%) and participant-oriented bundles (17%). Among the expert corpus, the proportion of text-oriented corpora is the highest. This indicates that experts place greater emphasis on using corpora to construct the structure of the text and to establish the logical connection of the argument. For example, experts frequently use resultative bundles such as "the results indicate that" and "the findings suggest that," as well as transition markers such as "on the other hand" and "in contrast," reflecting a stronger discourse organizational awareness. The highest proportion of research-oriented bundles in the master's corpus reflects master's students' tendency to describe research objects, methods, and procedures in detail in their abstracts, as evidenced by the frequent occurrence of bundles with specific research backgrounds such as "in the experimental class" and "the current situation of." The proportions of participant-oriented bundles are low in both corpora, which is consistent with the requirement of the abstract genre to maintain a relatively objective tone.

Regarding the specific use of research-oriented bundles, experts tend to use specific and precise quantification bundles, such as "90 participants" and "three groups of learners," whereas master's students prefer vague quantification bundles, such as "a great number of," "the total number of," and "the amount of." This difference reflects experts' stronger awareness of precise presentation of research data. In terms of procedure bundles, experts predominantly use passive voice constructions, such as "was measured using" and "was evaluated by," which align with the objectivity requirements of academic writing. In contrast, master's students mix passive voice with first-person constructions, such as "the author conducted," indicating an unstable awareness of academic register. Furthermore, master's students frequently use description bundles like "the current situation of," which is related to the applied nature of their research, as most master's theses aim to address practical problems in teaching and therefore need to describe the current situation as research background.

Regarding the specific use of text-oriented bundles, the most significant difference lies in the use of resultative markers. Experts frequently use hedging verbs such as "indicate" and "suggest" to express caution regarding research findings, whereas master's students predominantly use certainty verbs such as "show" and "conclude," lacking the necessary caution and reservation expected in academic writing. This finding is consistent with previous research showing that Chinese English learners tend to overuse boosters and underuse hedges[4]. In terms of transition markers, master's students show a marked preference for sequential markers such as "firstly...secondly...thirdly," reflecting a formulaic writing tendency,

ISSN: 2710-0170

whereas experts employ more diverse and natural transition devices. Regarding framing markers, master's students frequently use "according to" and "from the perspective of" to introduce theories or viewpoints, demonstrating a strong citation awareness, while experts more often use neutral framing markers such as "in the context of" and "based on," focusing more on situating their research within a broader academic context.

Regarding the specific use of participant-oriented bundles, the most prominent difference is in person choice. Expert writers commonly use the first-person plural "we," as in "we argue that," "we suggest that," and "we advocate for," establishing an interactive relationship with readers while expressing academic stance. In contrast, master's students rarely use the first person, instead using impersonal constructions such as "the author believes that" or "it can be seen that," appearing passive and detached. This difference may stem from an overemphasis on objectivity in academic writing instruction, leading students to mistakenly believe that first-person pronouns should be completely avoided, while neglecting the important role of the first person in establishing authorial authority and promoting interaction in academic discourse. Another significant difference is in the choice of modal verbs. Master's students' participant-oriented bundles almost always include "can," as in "it can be seen that" and "it can be found that," whereas experts more frequently use "should," as in "it should be noted that." "It can be seen that" implies the author stepping back from the text, while "it should be noted that" involves the author actively guiding the reader. This difference again confirms the characteristic of "author absence" in master's students' academic writing.

Synthesizing the above analyses, the main weaknesses in the use of lexical bundles in Chinese master's thesis abstracts are as follows: insufficient academic sophistication, with master's students preferring verb-based bundles and colloquial expressions while experts use more noun-based bundles and formal expressions; excessive certainty and lack of academic caution, with master's students almost exclusively using certainty verbs while experts use a large number of hedging expressions; author invisibility and lack of academic interaction, with master's students avoiding first-person pronouns and relying on impersonal constructions; and formulaic and narrow use of bundles, with master's students relying on only a few fixed expressions for transition markers, engagement markers, and quantifiers.

4. Conclusion

Based on two self-constructed corpora, this study has compared and analyzed the similarities and differences in lexical bundle use between the abstracts of Chinese master's theses and those of research articles by English expert writers. The main findings are as follows: in terms of overall distribution, the normalized frequency of lexical bundles in the expert abstracts is significantly higher than that in the master's abstracts, indicating that experts make more effective use of bundles to convey information efficiently within the limited space of an abstract; in terms of functional distribution, text-oriented bundles account for the highest proportion in the expert abstracts, reflecting stronger discourse organizational awareness, while research-oriented bundles account for the highest proportion in the master's abstracts, reflecting a focus on research methods and background details; in terms of specific use, the master's abstracts exhibit weaknesses such as a preference for verb-based bundles and vague expressions, excessive certainty and lack of academic caution, author invisibility and lack of academic interaction, and formulaic and narrow use of bundle patterns. This study has the following implications for the teaching of academic English writing: teachers should cultivate students' awareness of lexical bundles, guiding them to pay attention to frequent bundles in academic discourse rather than only individual words; teachers should help students distinguish between colloquial and academic bundles, avoiding the use of expressions with colloquial features in academic writing; teachers should guide students to appropriately use hedges such

ISSN: 2710-0170

as suggest and indicate, avoiding overly absolute expressions; teachers should correct the misconception that first-person pronouns must be completely avoided in academic writing and guide students to appropriately use the first person to express academic stance and establish interaction with readers. This study has the following limitations: the small corpus size may not fully represent the overall characteristics of the two author groups; the low minimum frequency threshold for bundle identification may have omitted some low-frequency but valuable bundles; and functional annotation involves some subjectivity. Future research could expand corpus size, incorporate more statistical testing methods, and attempt to combine interviews and other methods to gain deeper insights into the cognitive processes of master's students' bundle use. Further exploration could also investigate the relationship between bundle use and thesis quality to provide more targeted guidance for academic English teaching [5].

References

[1] Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgård & S. Oksefjell (Eds.), Out of corpora: Studies in honor of Stig Johansson (pp. 181-190). Rodopi.

[2] Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. English for Specific Purposes, 27, 4-21.

[3] Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. International Journal of Applied Linguistics, 18, 41-62.

[4] Hyland, K., & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. Journal of Second Language Writing, 6, 183-205.

[5] Xu, F. (2012). A study on the use and development features of English academic lexical bundles by Chinese learners. Foreign Languages in China, 9(04), 51-56.