

Article

Lexical Usage Features in International Chinese Language Education Academic Texts

Ze Yuan

Peking University, China

Received: 28 September, 2023/Accepted: 19 February, 2024/Published: 8 May, 2024

Abstract

This study conducted quantitative and statistical research on a self-built academic corpus to investigate the lexical usage features of academic texts in the field of International Chinese Language Education. The research found that the academic texts of International Chinese Language Education show a tendency towards popularization in vocabulary use. By analyzing the intersection of the general core words and academic words, it is found that such words have changed in meaning and function after entering the academic register, and should be treated as difficult points in the teaching of academic Chinese vocabulary.

Keywords

Academic Chinese, lexical usage features, academic vocabulary, corpus, popularization

1. Introduction

In recent years, the demand for international students to use Chinese for professional learning and academic research in China has been increasing. Proficiency in academic Chinese has become one of the educational objectives for international students (Gao & Liu, 2016). Research on the use of academic Chinese by international students (Zhu, 2014; Ji, 2015; Ji, 2016, etc.) and investigations into the academic Chinese writing needs of international students (Gao & Li, 2018; Gao, 2020; Li et al., 2020; Zhou, 2021, etc.) have revealed the following: Firstly, there is a significant gap between international students and native speakers in academic Chinese writing. Secondly, enhancing academic Chinese writing skills has become a crucial demand for international students studying in China.

Concerning the improvement of academic language proficiency in second language learners, the success or failure of learning academic vocabulary directly influences learners' reading abilities and academic achievements (Liu & Lei, 2020). The academic community has explored essential issues related to academic vocabulary, including conceptual classification, contextual functions, usage characteristics, skill development, and the creation of academic word lists. This study, based on a self-constructed corpus, aims to extract and conduct a detailed analytical study of high-frequency words appearing in academic texts in the field of international Chinese education. Theoretical aspects of

the research intend to address current focal questions in the field of academic vocabulary, such as the relationship between domain-specific academic vocabulary and general core vocabulary, and the domain characteristics of academic vocabulary. On a practical level, the study aims to assist research supervisors in this field to develop targeted vocabulary teaching plans, provide effective guidance to Chinese as a second language learners, and improve the appropriateness and accuracy of academic writing at the vocabulary level.

2. Literature Review

In terms of research on academic English vocabulary, researchers present the form and distribution of academic English at the lexical and phrase levels through empirical descriptions based on a large number of authentic language examples in corpora (Jiang, 2019). Specific research focuses on the conceptual classification of academic vocabulary (Paquot, 2010; Nation, 2013, etc.), functional framework and usage features analysis (Biber et al., 2004; Hyland, 2008, etc.), and the built and value discussion of academic word lists (Coxhead, 2000; Gardner & Davies, 2014; Lei & Liu, 2016, etc.).

As for Chinese academic vocabulary, research mainly focuses on the feature analysis of specific vocabulary items or linguistic markers in academic texts (Wu & Pan, 2010; Liu et al., 2021; Zhu & Hu, 2021). Zhang et al. (2020) explored the usage features of academic Chinese vocabulary. Based on corpus statistics, this study compared and analyzed the differences in vocabulary distribution frequency between natural science academic discourse and humanities and social science academic discourse. It focused on high-frequency function words, extracted a list of high-frequency function words in academic Chinese, and provided valuable teaching content references for advanced Chinese international language education, illustrating the usage features of high-frequency words in interdisciplinary academic texts in detail.

However, considering that the frequency, distribution, collocation, and meaning of the same vocabulary vary across different academic disciplines, some researchers argue that the so-called “*general academic vocabulary*” actually has limited benefits for learners (Hyland & Tse, 2007; Durrant, 2016). Reflected at the lexical level, for example, the word “*process*” (meaning “procedure”) and its homonym “*process*” (meaning “to treat materials”) are both common in engineering, natural sciences, and social sciences corpora, but natural sciences and engineering disciplines tend to use it more as a noun than social sciences. Therefore, the best way to prepare students for learning is not to search for universally applicable (lexical) teaching materials, but to understand the discourse characteristics of specific fields. It is recommended that researchers establish more targeted vocabulary databases based on a single discipline and conduct more targeted research (Hyland & Tse, 2007).

Taking a comprehensive view, current research of academic Chinese lacks a focused investigation into academic vocabulary within specific and singular academic domains. The field of International Chinese Language Education is currently an emerging interdisciplinary subject, and its research outcomes have significant guiding implications for language teaching practices among international students. Concurrently, an increasing number of international students are choosing this field as their academic major. Therefore, researching the features of vocabulary usage in this field can assist research supervisors in this field to develop more targeted vocabulary teaching plans and provide effective guidance to Chinese as a second language learners.

The specific research questions are as follows:

1. What are the overall usage features of vocabulary in International Chinese Language Education academic texts?
2. What are the usage features of high-frequency word in this field’s academic texts?
3. From the perspective of lexical usage features, what recommendations can be made for teaching academic Chinese vocabulary?

From a macro perspective, Question 1 investigates the characteristics of vocabulary usage in this field, while Question 2 delves into the specifics from a relatively micro viewpoint. By combining both, we can obtain a relatively comprehensive understanding of the vocabulary usage features in this field. The conclusions drawn from Questions 1 and 2 serve as the basis for addressing Question 3.

3. Research Methods

Based on our self-built academic text corpus, we will extract high-frequency words, academic words, and the general core words from academic texts in the field of International Chinese Language Education, and conduct detailed analytical research on them.

3.1 Corpus building

We conducted equidistant sampling of articles from five types of academic journals in the field of International Chinese Language Education, including “*Language Teaching and Linguistic Studies*”, “*Chinese Teaching in the World*”, “*Chinese Language Learning*”, “*Applied Linguistics*” and “*TCSOL Studies*”. We established a corpus of 3 million Chinese characters.

These five types of journals are high-quality journals in the field, ensuring the standardization and accuracy of vocabulary usage. We categorized the articles based on several main research directions in this field and created four sub-corpora¹: Corpus of Chinese Ontology Theory and Application Research, Corpus of Chinese Acquisition and Testing Research, Corpus of Chinese Teacher Development Research, and Corpus of Chinese Teaching and Curriculum Research. The specific distribution and capacity of the corpora are shown in Table 1.

Table 1

Distribution of International Chinese Language Education Academic Sub-Corpora

Corpus Name	Corpus of Chinese Ontology Theory and Application Research	Corpus of Chinese Acquisition and Testing Research	Corpus of Chinese Teacher Development Research	Corpus of Chinese Teaching and Curriculum Research
Corpus Size	793,023 characters	768500 characters	630755 characters	815106 characters

3.2 Word segmentation processing

In this study, we employed the Chinese word segmentation toolkit, *pkuseg*, developed and released by the Language Computing and Machine Learning Research Group at Peking University, for word segmentation processing. This segmentation method demonstrates high accuracy in word segmentation across multiple domains (Zhang & Fan, 2022). The *pkuseg* toolkit maintains a word segmentation accuracy of over 95% for a portion of the test corpus. After manual verification and correction of individual segmentation errors, a total of 1,983,655 word tokens and 63,635 word types were obtained.

3.3 High-frequency word extraction

To extract high-frequency words from International Chinese Language Education texts, we follow the mainstream standards used in building word lists (Coxhead, 2000; Gardner & Davies, 2014; Lei & Liu, 2016), employing the criteria of “*range*” and “*minimum frequency*”. It should be noted that this study focuses on the usage features of vocabulary that frequently appears in academic texts in this field, not

limited to academic vocabulary. Therefore, the standards commonly used in building academic word lists, such as “*specialization*” (Coxhead, 2000; Zhao, 2015; Wang & Wang, 2022), “*ratio*” (Gardner & Davies, 2014; Lei & Liu, 2016), and “*dispersion*” (Gardner & Davies, 2014; Wang & Wang, 2022), are not used as criteria for extracting high-frequency words.

(1) **Range:** High-frequency words in academic texts in this field should be distributed across three or more of the four sub-corpora.

(2) **Minimum Frequency:** Referring to the frequency standards used in the creation of current international academic word lists (Coxhead, 2000; Lei & Liu, 2016), we stipulate that the frequency of word occurrences should be 30 times per million words or more.

After completing the initial vocabulary screening, we proceed with manual proofreading and verification. Human intervention is primarily reflected in the following three aspects:

Firstly, for high-frequency words to be analyzed in academic texts in this field, we strictly refer to the “*Contemporary Chinese Dictionary*” (7th edition), excluding high-frequency phrases (such as “*Zuo chu*”[make out] and “*Kan dao*”[have seen]) and proper nouns (personal names) that appear in machine word segmentation.

Secondly, there are also some specialized academic terms that are not included in the “*Contemporary Chinese Dictionary*” (7th edition), such as “*Gou shi*”(linguistic terminology which means construction) and “*Lun yuan*” (linguistic terminology which means argument). However, as scientific vocabulary with specialized semantics and semantic invariance, and lacking precise synonyms, they are equally important for academic writing. In this study, we retain them under the premise of meeting the extraction criteria mentioned above.

Thirdly, part-of-speech restriction: Zhang et al. (2020) pointed out that the quantity of function words such as conjunctions, prepositions, and adverbs is relatively limited. By extracting and comparing high-frequency function words in different disciplines, core function words across disciplines can be selected. These core function words can be regarded as the core vocabulary of general academic Chinese. From a practical perspective, these function words play a role in text organization, discourse coherence, and are the key and difficulty for Chinese as a second language learners to improve the appropriateness of academic writing. Based on Zhang et al.’s (2020) research, we also consider nouns, verbs, adjectives, conjunctions, prepositions, and adverbs appearing in academic texts in this field as objects for statistical analysis.

3.4 Typical academic words extraction

For the extraction of academic terms in academic texts in this field, we consider the vocabulary included in our self-built “*International Chinese Language Education Academic Vocabulary List*” as typical academic terms in this field. This vocabulary list is created based on mainstream paradigms, through the process of building an academic text corpus, word segmentation processing, setting selection criteria, and manual intervention. The corpus and segmentation tool used to generate the vocabulary list are the same as those used in this study. The vocabulary list contains a total of 751 terms, covering 15.8% of academic texts in this field, which is much higher than its coverage in non-academic text corpora (2.4%). The included academic words have strong representativeness. The selection criteria are as follows:

(1) **Range:** The vocabulary included in the list should be distributed in at least three out of the four sub-corpora.

(2) **Frequency:** The frequency of vocabulary occurrence should be 30 times per million words or more.

(3) **Dispersion:** The minimum threshold for the dispersion of included terms is 0.30.

(4) **Ratio:** The frequency of occurrence in the academic text corpus should be 1.5 times or more than that in the larger non-academic text corpus.

In addition, to maintain consistency with the high-frequency term analysis, we also exclude high-frequency phrases that are allowed in the academic vocabulary list, as well as words other than nouns, verbs, adjectives, conjunctions, prepositions, and adverbs. After the screening process, a total of 724 academic terms are obtained.

3.5 Extraction of general core words in contemporary Chinese

Zhang (2006) pointed out that after decades of teaching and repeated statistical analysis of various language materials in the field of teaching Chinese as a foreign language, a consensus has been reached on the vocabulary size, namely 3000 words, 5000 words, and 8000 words can be used as three benchmarks for vocabulary teaching. Therefore, we take the “*Lexicon of Common Words in Contemporary Chinese*” (Research Group of the Lexicon of Common Words in Contemporary Chinese, 2008)² with word frequency rankings from 1 to 3000 as the standard for general core words.

3.6 Analysis of word usage features

Macro-Perspective: It is mainly reflected in two aspects: the cumulative frequency of vocabulary and vocabulary density. The cumulative frequency of vocabulary is one of the important criteria for judging the conservatism of vocabulary selection in a certain type of text, which can be determined by the proportion of word types in the total number of words. Vocabulary density refers to the proportion of content words to the total number of words in the text. Vocabulary density affects the perceived difficulty of readers in understanding the text content (Zhang et al., 2020).

Micro-Perspective: Usage rate and proportion of general core words and academic words in high-frequency words. By examining the usage rate and proportion of general core words and academic words in high-frequency words, we can understand the tendency of vocabulary usage in academic texts.

Analysis of vocabulary difficulties in academic Chinese teaching: We focus on the intersection of general core words and academic words in high-frequency words. We examine how the meaning and function of general core words change when they enter academic texts.

4. Research Findings

4.1 Macro-perspective

In terms of the cumulative frequency of vocabulary, after machine segmentation and manual proofreading and modification, we obtained a total of 1,983,655 word tokens and 63,635 word types in the corpus of International Chinese Language Education academic texts (excluding duplicate words shared among sub-corpora). The average cumulative frequency of vocabulary is 3.21%. As mentioned earlier, the cumulative frequency of vocabulary in an article typically represents the level of conservatism in its word choice. Generally, the higher the cumulative frequency of vocabulary, the more flexible the text’s word usage and the lower the level of conservatism. Please refer to Table 2 for detailed accumulated frequency of vocabulary in each sub-corpus.

Based on the situation of various sub-corpus, the cumulative frequencies of vocabulary in different directions are arranged from high to low as follows: *ontology research* (5.37%) > *acquisition and testing research* (4.63%) > *teacher development research* (4.45%) > *curriculum and teaching research* (4.36%).

Table 2

Cumulative Frequency of Vocabulary in the International Chinese Language Education Academic Corpus

Corpus	Corpus of Chinese Ontology Theory and Application Research	Corpus of Chinese Acquisition and Testing Research	Corpus of Chinese Teacher Development Research	Corpus of Chinese Teaching and Curriculum Research	Total
Token	549,072	510,208	398,251	526,124	1,983,655
Type	29,506	23,280	17,741	22,959	63,635
Cumulative Frequency	5.37%	4.63%	4.45%	4.36%	3.21%

We also conducted a cross-sectional comparison of the cumulative frequency of vocabulary in the field of International Chinese Language Education. By comparing it with the cumulative frequency of vocabulary in general academic and non-academic texts, we aim to provide a more scientific explanation for the conservative use of academic vocabulary in this field. For the cumulative frequency of vocabulary in general academic texts, we utilized a self-built general academic text corpus covering 12 subfields, including History, Sociology, Chinese Literature, Law, Management, Education, Earth Science, Ecology, Physics, Mechanical Engineering, Computer Science, and Information and Communication Engineering. The corpus consisted of 585 specialized journal articles, which, after word segmentation using the *pkuseg*, yielded 3,340,476 word tokens and 128,588 word types. The cumulative frequency of vocabulary in general academic texts was found to be 3.85%. As for the cumulative frequency of vocabulary in non-academic texts, we used a self-built non-academic text corpus, which included contemporary novels, news articles, and encyclopedic texts. After word segmentation using the *pkuseg*, the corpus yielded 12,613,607 word tokens and 394,573 word types. The cumulative frequency of vocabulary in non-academic texts was found to be 3.13%. Detailed information on the cumulative frequency of vocabulary in the three different types of texts can be found in Table 3.

Table 3

Cumulative Frequency of Vocabulary in Three Different Types of Corpora

Corpus	Token	Type	Cumulative Frequency
Corpus of the International Chinese Language Education	1983655	63635	3.21%
Corpus of General Academic Text	3340476	128588	3.85%
Corpus of Non-Academic Text	12613607	394573	3.13%

The second quantitative indicator is vocabulary density. Vocabulary density refers to the proportion of content words in the total number of words in a text. It is generally believed that the higher the vocabulary density in a text, the more difficult it is for readers to comprehend. After machine segmentation and part-of-speech tagging, the quantity of content words in the *International Chinese Language Education Academic Corpus* is 1,425,117, accounting for 71.8% of the total word count. In the *General Chinese Academic Text Corpus*, the frequency of content words is 2,385,520, accounting for 71.4% of the total word count. In the *Non-Academic Text Corpus*, the frequency of content words is 10,647,806, accounting for 84.4% of the total word count.

4.2 Micro-perspective

According to the research methodology mentioned above, after machine segmentation, setting selection criteria, machine screening, and manual verification, we obtained a total of 2,099 high-frequency words in academic texts of International Chinese Language Education. The total word frequency of these 2,099 words is 1,042,265, which accounts for 54.5% coverage of the corpus of International Chinese Language Education academic texts. Among them, there are a total of 724 academic words, accounting for 34.5% of the total high-frequency words. There are a total of 1,172 general core words in contemporary Chinese, accounting for 55.8%. Meanwhile, there are 133 vocabulary items that belong to both academic words and general core words, accounting for 6.3%. The distribution of different vocabulary categories in the high-frequency words is detailed in Figure 1.

Figure 1

Distribution of Academic Words and General Core Words in High-frequency Vocabulary

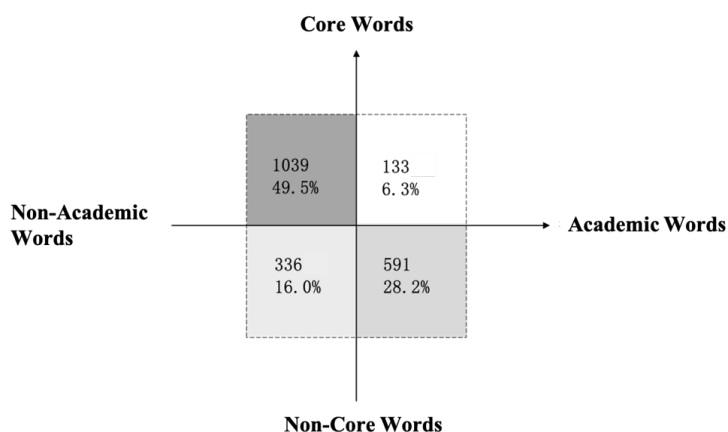
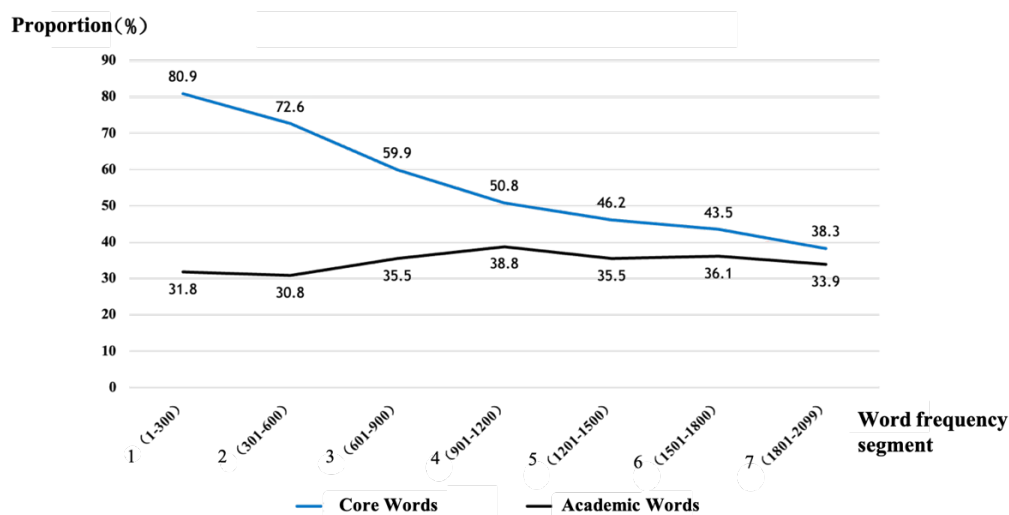


Figure 2

Proportion of Academic Words and General Core Words in Different Frequency Ranges



In order to further investigate the usage of academic words and general core words in high-frequency words, we divided 2099 high-frequency words into seven segments according to their frequency index³, and then calculated the proportion and trend of academic terms and general core words in each frequency segment. The statistics reveal that general core words have a relatively high proportion in the first two

frequency segments, accounting for 80.9% and 72.6% respectively. The proportion of general core words shows a decreasing trend as the frequency decreases, with the lowest proportion observed in the seventh frequency segment (38.3%). The variation trend of academic terms in different frequency segments is not significant, with the highest proportion observed in the fourth frequency segment (38.8%) and the lowest proportion in the first frequency segment (31.8%). For more details on the proportion and trend of academic terms and general core words in different frequency segments, please refer to Figure 2.

As mentioned earlier, we focus on examining the usage of six classes of words in this academic text on International Chinese Language Education: nouns, verbs, adjectives, adverbs, conjunctions, and prepositions. After segmenting and part-of-speech tagging using the *pkuseg*⁴, as well as conducting frequency analysis, we obtained the “*Distribution of Word Classes for High-frequency Words in Academic Texts on International Chinese Language Education*”, as shown in Table 4.

Table 4

Distribution of Word Classes in High-frequency Words in International Chinese Language Education Academic Texts

	Noun	Verb	Adjective	Adverb	Preposition	Conjunction
Quantity	904	797	182	122	34	72
Proportion	43.1%	40.0%	8.7%	5.8%	1.6%	3.4%

5. Discussion

5.1 A tendency towards popularization in vocabulary use

From the perspective of lexical features, we first examined the cumulative frequency and lexical density of vocabulary in International Chinese Language Education academic texts from a macro perspective, and compared them with our self-built *general academic Chinese corpus* and *non-academic text corpus*. The results showed that the highest cumulative frequency was found in the general academic Chinese corpus (3.85%), followed by the International Chinese Language Education academic corpus (3.21%), and the non-academic text corpus had the lowest cumulative frequency of vocabulary (3.13%). Cumulative frequency of vocabulary is an important criterion for judging the conservatism of vocabulary selection in a specific linguistic style. Zhang et al. (2020) examined the usage features of general academic Chinese texts and found an average cumulative frequency of vocabulary of 3.9%. At the same time, he cited the research of Lue & Yao (2018), which pointed out that the cumulative frequency of vocabulary in Laws texts is about 3.5%, and in novel texts it is about 10.6%. Therefore, it can be concluded that the academic register exhibits a higher degree of conservatism in vocabulary selection, adhering to established norms and showing minimal use of neologisms.

However, through cross-comparison, we find that the conservative degree of vocabulary usage in the International Chinese Language Education academic register is similar to that of previous studies on Laws texts (3.5%), but lower than that of previous studies on novel texts (10.6%). More importantly, the cumulative frequency of academic vocabulary in this field (3.21%) is very close to that of our self-built *non-academic text corpus* (including contemporary novels, press releases, and encyclopedic knowledge) (3.13%). Therefore, it seems that we cannot conclude that the conservative degree of vocabulary usage in academic texts in this field is significantly higher than that of other genre types (excluding novels). On the contrary, in terms of the richness of vocabulary choices, academic texts in International Chinese Language Education approach other genre types and do not exhibit significant conservatism. The density of vocabulary also confirms the above conclusion. The density of vocabulary affects the perceived difficulty of readers in understanding the content of a text. Texts with high vocabulary density contain

a wide range of information domains, and people find it more difficult to perceive texts with high vocabulary density compared to texts with low vocabulary density (Lue & Yao, 2018). However, through statistics, we find that the vocabulary density (71.8%) of academic texts in this field is similar to that of general academic texts (71.4%), and lower than the vocabulary density of the *non-academic text corpus* (84.4%). Therefore, from the perspective of macro-level vocabulary usage features, academic texts in International Chinese Language Education do not exhibit prominent features of low vocabulary choice richness and high perceived difficulty for readers, which are close to the vocabulary selection profile of other genre types.

From a more micro perspective, using the operational definition mentioned earlier, we identified 2,099 high-frequency words in academic texts in this field, with an overall coverage rate of 54.5%. Among these words, there were 724 academic words and 1,172 words from the general core words. In these high-frequency words, the majority (55.8%) were still from the general core words. As shown in *Figure 1*, the most common type of high-frequency words is those that belonged to the general core words but were non-academic words (49.5%). This indicates that among the high-frequency words with high academic coverage in this field, the most frequently used words are still from the general core vocabulary of contemporary Chinese. In addition, in terms of the proportion and trend of general core words and academic words in different frequency ranges, the general core words had an absolute advantage in the high-frequency range (80.9%, 72.6%), and its proportion gradually decreased as the frequency range decreased, while the proportion of academic words remained consistently low. Finally, we extracted the top ten members from the high-frequency words in terms of adverbs, prepositions, and conjunctions. The statistics revealed that out of the 30 words, 29 of them belong to the general core words of contemporary Chinese, while one word “Yu”(and) does not. None of the 30 words belong to the international academic vocabulary of Chinese language education.

From a more microscopic perspective, it also shows that academic texts in this field tend to use general core words, presenting a tendency towards popularization in lexical usage. In terms of the essential function of academic texts in this field, it is to convey the author’s viewpoints and findings to the readers through rigorous argumentation. Therefore, it is necessary to ensure the accessibility and comprehensibility of the vocabulary used in order to effectively convey the intended meaning.

However, while the academic vocabulary in this field tends to have popularized features, academic papers still pose a certain level of difficulty for readers, reflecting the inherent rigor and specialization of academic discourse. Apart from the reading barriers caused by specialized academic vocabulary, it may be due to the fact that the stylistic features of academic discourse are largely achieved through academic phrases (lexical bundles) or fixed structures that serve specific functions (Biber et al., 2004; Hyland, 2008; Simpson-Vlach & Ellis, 2010). This view is also reflected in the research on Chinese academic discourse. Ji (2016) examined the use of classical Chinese structures in five categories of Chinese academic papers written by Korean students: “Yu” (“in”), “Zhe” (“a third-person pronoun”), “Zhi” (“of”), “Er” (“and”), and “Yi” (“by”). Among them, three of these items ranked among the top ten in terms of specific word categories in International Chinese Language Education academic texts (*Yu*, *Er*, *Yi*). The study also indicates that there is a reasonable level of proficiency in terms of vocabulary, particularly in the specialized terminology, but there is still a significant gap at the level of grammar structure.

Based on this, we propose two suggestions for guiding second language learners in academic writing. Firstly, the extraction of high-frequency academic vocabulary based on the frequency-driven concept is an effective approach in teaching academic vocabulary. Vocabulary instruction should first help learners master correct and standard high-frequency vocabulary, without excessively pursuing word complexity or vocabulary difficulty. Secondly, after acquiring the semantic and usage of high-frequency and academic words, learners should gradually transition to the study of more complex lexical chunks, grammar structures, and other modules in order to enhance the appropriateness of academic writing and highlight the genre features of academic texts.

5.2 The influence of academic register on vocabulary usage

Halliday & Hasan (1985) define register as “a variety of language that is associated with a particular social context and can be understood within that context” (Feng, 2006). Academic language, as a specialized register, has certain impacts on the meaning and function of its lexical members. Some academic words have meanings that differ from their commonly used counterparts, and the meanings of certain academic terms are rarely seen in general language (Zhang, 2022). The influence of register on vocabulary is often a challenge for learners. Bloor & Bloor (1986) proposed the “*Common Core Hypothesis*,” suggesting that every register has a dominant set of syntactic and lexical core items, and learners should first acquire these basic linguistic components that constitute the core before focusing on the specific target language of a particular discipline. If learners first acquire the common meanings or conventional usages, they are more likely to misuse these in a different register, requiring a process of relearning.

In this study, we found that the academic register of this field has a significant impact on lexical usage, mainly concentrated in the first quadrant as shown in Figure 1, which represents the intersection between the general core words and academic words. There are a total of 133 words in this section, accounting for 6.3% of the overall high-frequency words. We provide examples to illustrate this (see Table 5).

Table 5

Differences between General and Academic Meanings of High-frequency Words in International Chinese Language Education Academic Texts

Word	general meaning	academic meaning
Jia gong	1. To transform raw materials, semi-finished products, etc. into finished products, or to bring them up to specified requirements; 2. To carry out various activities aimed at making the finished product more perfect and refined.	Processing activities of the brain in response to input stimuli.
Ting shuo	hear of	Listening and Oral Expression
He fa	In compliance with legal regulations.	Complies with grammatical rules.

Regarding such vocabulary, we recommend that teachers enhance the distinction between its academic and general meanings. They can utilize the contexts in which the vocabulary appears in academic discourse to emphasize its academic meaning and usage.

6. Conclusion

This study conducted quantitative and statistical research on a self-built academic corpus to investigate the lexical usage features of academic texts in the field of International Chinese Language Education. A comparison was made with the self-built “*Corpus of General Academic Text*” and “*Corpus of Non-Academic Text*” to analyze the usage features of high-frequency words in this field. The study found that academic texts in the field of International Chinese Language Education did not exhibit excessive conservatism or high lexical density in terms of vocabulary selection. Among the 2099 high-frequency words, the usage of the general core words in standard contemporary Chinese dominated. Based on the research findings, it is concluded that academic texts in International Chinese Language Education demonstrate features of popularization in vocabulary usage.

At the same time, the academic register has an impact on the meaning and function of the general core words. Although the intersection between the general core words and academic words is relatively small, learners need to acquire some of these words as part of their academic vocabulary learning, which should be treated as a difficulty in academic vocabulary teaching. In order to systematically study the semantic and functional changes of general academic vocabulary in academic texts, further detailed and comprehensive research is needed.

Notes

1. The sub-corpus is categorized based on various sub-disciplines within the field of International Chinese Language Education, as well as different research directions established by certain Chinese universities.
2. The specification (draft) presents a list of 56,008 commonly used Mandarin Chinese words in contemporary social life, which are relatively stable and frequently used. The specification also provides the word forms of these words. It can be used as a reference for primary and secondary school Chinese language teaching, literacy education, Chinese language education, Chinese information processing, and dictionary compilation.
3. The frequency index is arranged in ascending order from 1 to 2099, with a lower frequency index indicating a higher occurrence of the word in the International Chinese Language Education Academic Corpus.
4. There are cases of ambiguous word class in part-of-speech tagging. In such cases, when conducting frequency statistics, we count the word in multiple word classes simultaneously. For example, the word “Kun nan” (difficulty) can function as both a noun and an adjective. We count this word in both the noun and adjective categories.

References

- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied linguistics*, 25(3): 371-405.
- Bloor, M., & Bloor, T. (1986). *Languages for specific purposes: Practice and theory*. CLCS Occasional Papers, Trinity College.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Douglas, B., Susan, C., & Viviana, C. (2004). If you look at ...: lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405.
- Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing? *English for Specific Purposes*, 43, 49-61.
- Feng, Z X. (2006). 现代语言学流派概论 [Introduction to Modern Linguistic Schools], 北京大学出版社 [Peking University Press].
- Gao, Z X., & Li, S. (2018). 学术汉语写作教材建设刍议 [On the complication of textbooks for Chinese academic writing]. 云南师范大学学报(对外汉语教学与研究版) [*Journal of Yunnan Normal University(Teaching & Studying Chinese as a Foreign Language Edition)*], 16(6), 12-21.
- Gao, Z X., & Liu, F Y. (2016). 论学术汉语在对外汉语教学中的重要性 [On the role of the CAP teaching in Teaching Chinese to Speakers of Other Language], 云南师范大学学报(对外汉语教学与研究版) [*Journal of Yunnan Normal University(Teaching & Studying Chinese as a Foreign Language Edition)*], 14(2), 44-5.

- Gao, Z X. (2020). 留学生研究生汉语学术论文写作需求及能力调查 [A needs analysis of the skills of international postgraduates' academic writing in Chinese], 云南师范大学学报(对外汉语教学与研究版) [*Journal of Yunnan Normal University(Teaching & Studying Chinese as a Foreign Language Edition)*], 18(6), 68-75. <https://doi:10.16802/j.cnki.ynsddw.2020.06.010>.
- Gardner, D., & Davies, M. (2014). A New Academic Vocabulary List. *Applied Linguistics*, 35(3), 305-327.
- Halliday, M A K., & Hasan, R. (1985). *Language, context, and text: Aspects of language in a social-semiotic perspective*. Oxford University Press.
- Hyland, K., & Tse, P. (2007). Is there an "Academic Vocabulary"? *TESOL Quarterly*, 41(2), 235-253.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21.
- Ji, C B. (2016). 韩国学生汉语学术论文中文言结构使用初探 [Exploration of the Usage by Korean Students of the Classical Chinese Structures in Academic Paper], 汉语学习 [*Chinese Language Learning*], 6, 77-85.
- Ji, Y Q. (2015). 话语标记语在留学生汉语学术写作中的使用研究 [Studies on the use of discourse markers in the Chinese academic writing of foreign students], 南京大学硕士学位论文 [Master Thesis of Nanjing University].
- Jiang F. (2019). 语料库与学术英语研究 [Corpora and EAP studies], 外语教学与研究出版社 [Foreign Language Teaching and Research Press].
- Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, 22, 42-53.
- Li, H Y., Zhang, W X., & Xin, P. (2020). 本科留学生学术汉语写作课需求调查与课程建设——以北京大学本科留学生学术汉语写作通选课为例 [The need assessment and curriculum development of Chinese academic writing for international undergraduate students—Take Chinese academic writing as general selective course for international undergraduate students in Peking University as an example], 国际汉语教育(中英文) [*International Chinese Language Education*], 5(1), 51-60.
- Liu, D L., & Lei L. (2020). 学术词表研究综述 [A review of the study on academic vocabulary list]. 外语教学 [*Foreign Language Education*], 41(2), 34-38+50.
- Liu, R., Huang, Q Q., & Wang S. (2021) 汉语学术语篇转述标记的形式、功能与分布 [The Form, Function and Distribution of Reporting Markers in Chinese Academic Papers], 当代修辞学 [*Contemporary Rhetoric*], 6, 60-72.
- Lv, W T., & Yao S Y. (2018). 词汇规制与立法语言的简明性 [Vocabulary regulation and the simplicity of legal language], 语言文字应用 [*Applied Linguistics*], 4, 65-74.
- Nation, I., & Hunston, S. (2013). *Learning vocabulary in another language*. Cambridge University Press.
- Paquot, M. (2010). *Academic vocabulary in learner writing: From extraction to analysis*. Continuum.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512.
- Wang, X R., & Wang, J M. (2022). 经贸类本科专业学术汉语词表研究 [Establishment of an academic word list for economics and trade majors], 语言教学与研究 [*Language Teaching and Linguistic Studies*], 4, 9-19.
- Wu, G Q., & Pan, C L. (2010). 汉语学术论文中作者立场标记语研究 [Authorial Stance Markers in Chinese research articles], 语言教学与研究 [*Language Teaching and Linguistic Studies*], 3, 91-96.
- Zhang, B. (2022). 学术汉语词汇的主要特点及教学策略 [Linguistic features and teaching strategies of Chinese academic vocabulary], 世界汉语教学 [*Chinese Teaching in the World*], 36(4), 517-530.

- Zhang, C., Li, J P., & Shen S X. (2020). 学术汉语的词汇使用特征研究 [A study on the lexical usage characteristics of academic Chinese], 语言教学与研究 [Language Teaching and Linguistic Studies], 6, 19-27.
- Zhang, H S. (2006). 外国学生汉语词汇学习状况计量研究 [A quantitative study of vocabulary development of foreign students learning Chinese as a foreign language], 世界汉语教学 [Chinese Teaching in the World], 1, 70-76+3.
- Zhang, Q., & Fan, Y S. (2022). 基于改进T5 PEGASUS 模型的新闻文本摘要生成研究 [Research on generating news texts summarization based on improved T5 PEGASUS model], 电子科技 [Electronic Science and Technology], 36(12), 72-78.
- Zhao, Z G. (2015). 专门用途英语学术词表创建研究——以航海英语为例 [Development of ESP Academic Word Lists — A Case Study of Maritime English], 重庆交通大学学报（社会科学版） [Journal of Chongqing Jiaotong University(Social Sciences Edition)], 15(6), 140-144.
- Zhou, Z J. (2021). 留学生学术汉语课程需求调查与分析 [Investigation and needs analysis of Chinese for academic purpose course for foreign students], 云南师范大学硕士学位论文 [Master's thesis of Yunnan Normal University].
- Zhu, G Z. (2014). 留学生本科毕业论文中的语体特征和语体意识研究——以复旦大学留学生毕业论文为例 [A study of stylistic features and stylistic awareness in the undergraduate thesis of international students: taking the graduation thesis of international students at Fudan University as an example], 复旦大学硕士学位论文 [Master's thesis of Fudan University].
- Zhu, Y., & Hu, X D. (2021). 汉语连词在不同学术语域的聚合:多维度定量分析 [A multi-dimensional analysis of variations in the use of conjunctions in various registers of academic Chinese], 语言教学与研究 [Language Teaching and Linguistic Studies], 2, 57-69.

Ze Yuan is a PhD Candidate in the School of Chinese as A Second Language at the Peking University. His research interests include corpus linguistics and second language acquisition.

国际中文教育学术语篇词汇使用特征研究

袁泽

北京大学，中国

摘要

本研究通过对 300 万字的自建学术文本语料库进行量化统计研究，考察了国际中文教育领域学术文本的词汇使用特征，同时对自建“汉语通用学术文本语料库”“非学术文本语料库”中的词汇使用特征进行了横向比较。重点分析该领域学术本文中的高频词使用特征。研究发现，国际中文教育领域的学术文本在词汇选项保守程度、词汇密度方面，并未显示出过于保守、词汇密度较高等情况；在 2099 个高频词中，现代汉语通用核心词的使用占主导地位。根据研究结果，我们认为国际中文教育学术文本在词汇使用方面体现出平易化特征。同时，学术语域对于现代汉语通用核心词的词义和功能均有影响。根据共核假设理论，现代汉语通用核心词与学术词的交集部分虽占比较少，但因学习者需要对其中的部分词汇进行二次习得，该部分应作为学术词汇教学的难点来对待。

关键词

学术汉语，词汇使用特征，学术词汇，语料库，平易化

袁泽，北京大学在读博士研究生，研究方向：语料库语言学、二语习得。