Original Paper

Structural and Functional Usages of Lexical Bundle in Native and Non-Native English Sociology Academic Writings: A Comparative Study Based on MICUSP

Bingren Long[1], Hui Shao[2*], Juan Yao[3] & Junchen Zhang[4]

[1] College of Foreign Languages, Beijing University of Technology, Beijing, China

[2*] College of Foreign Languages, Beijing University of Technology, Beijing, China

[3] College of Foreign Languages, Beijing University of Technology, Beijing, China

[4] College of Foreign Languages, Beijing University of Technology, Beijing, China

Abstract

Framed within Biber's structural classification and Hyland's functional classification, this study investigates the structural and functional uses of four-word lexical bundles in sociology academic writings by native (NS) and non-native speakers (NNS) of English by using the Michigan Corpus of Upper-level Student Papers (MICUSP) and tools such as AntConc and Xu Jiajin's Log-Likelihood Ratio Test Calculator. The statistical findings reveal that both NS and NNS predominantly use noun phrase-based bundles, reflecting their critical role in sociology academic discourse. However, NNS exhibit a significantly higher usage of prepositional phrase-based bundles, indicating a strategy to enhance clarity and coherence. Functionally, while both groups rely heavily on research-oriented bundles, NNS employ text-oriented bundles more frequently, suggesting a heightened focus on ensuring textual coherence as a compensatory strategy for managing discourse. Conversely, NS more frequently use participant-oriented bundles, indicating greater rhetorical sophistication and personal engagement with the reader. The study contributes to linguistic theories by elucidating how lexical bundles function within the context of sociology and offering practical implications for English for Academic Purposes (EAP) instruction in sociology, emphasizing the need for targeted support in both the structural and

functional aspects of lexical bundles and in improving academic writing skills.

1. Introduction

The acquisition of fluent and idiomatic language is a key challenge for learners of any language, particularly in academic writing where precision and clarity are paramount (Hyland, 2008). This challenge is further amplified for non-native writers who must navigate the complexities of a new linguistic system while simultaneously mastering the conventions of academic discourse within specific disciplines (Biber, 2006). One area of growing interest in this domain is the study of multi-word units, particularly lexical bundles, and their role in characterizing fluent and idiomatic language production.

Lexical bundles are frequently occurring sequences of words that act as building blocks of fluent discourse (Biber, 2006). Understanding the nature and use of these bundles is crucial for gaining insights into the underlying linguistic features that distinguish expert from novice writing, and native from non-native production. Corpus linguistics, with its ability to analyze large amounts of language data, provides a powerful tool for investigating such patterns. Numerous studies have utilized corpus linguistic approaches to investigate lexical bundles across a variety of disciplines and registers. These studies have highlighted the significant role played by lexical bundles in achieving textual cohesion, establishing disciplinary conventions, and signaling rhetorical functions.

However, although numerous studies have explored the use of lexical bundles among native and non-native learners across various disciplines such as history, biology, and engineering, there is a noticeable scarcity of research specifically targeting sociology students. Moreover, most existing studies on lexical bundles have concentrated on specific sections of academic papers, such as titles, abstracts, introductions, and conclusions (Pan, 2016; Xu, 2014). Few studies provide a more holistic view of lexical bundle usages to reveal how these bundles contribute to the overall coherence and cohesion of academic writing across different sections.

This study aims to bridge identified research gaps by analyzing the use of four-word lexical bundles within the full texts of sociology academic papers in the Michigan Corpus of Upper-level Student Papers (MICUSP). MICUSP's diverse compilation of student writings across various disciplines and linguistic backgrounds offers a comprehensive dataset for contrasting the lexical practices of native speakers (NS) and non-native speakers (NNS). Utilizing AntConc, a powerful corpus analysis tool, this research identifies and extracts the most frequent four-word lexical bundles from the sociology sub-corpus of MICUSP. These bundles are subsequently analyzed through Biber's (2006) structural classification and Hyland's (2008) functional classification. The study also employs Xu Jiajin's Log-Likelihood Ratio Test Calculator to rigorously evaluate differences in lexical bundle usages between NS and NNS, thus laying a solid statistical foundation for this comparative analysis.

12

This study contributes to the understanding of 4-word lexical bundles in NS and NNS sociology academic writings by answering the following research questions:

(1) From the structural perspective, what are the similarities and differences in the usages of lexical bundles between NS writers and NNS writers in sociology academic papers?

(2) From the functional perspective, what are the similarities and differences in the usages of lexical bundles between NS writers and NNS writers in sociology academic papers?

(3) What are the potential reasons for those similarities and differences?

This study holds both theoretical and practical significance. Theoretically, it adds depth to existing linguistic theories on lexical bundles by elucidating how these elements function within the specific context of sociology. It provides empirical evidence on the structural and functional use of lexical bundles, thus refining our understanding of text cohesion and coherence in academic discourse. Practically, the findings from this study have direct implications for teaching academic writing, especially for non-native speakers, contributing to the broader field of English for Academic Purposes (EAP).

2. Literature Review

2.1 Definitions of Lexical Bundles

The idea of lexical bundles evolved from early observations about recurring patterns in language to a central concept in corpus linguistics. Linguists like Firth and Halliday first highlighted the significance of collocations and fixed expressions. However, it was the rise of corpus linguistics in the 1980s and 1990s, spearheaded by researchers like Sinclair and Stubbs, that enabled the systematic analysis of these patterns in vast amounts of real language data. This led to the identification of frequently recurring multi-word sequences, termed "lexical bundles" by Biber and colleagues in their influential 1999 work. The concept has since gained prominence, with research focusing on the functions, variations, and implications of these prefabricated units for language use and acquisition.

Lexical bundles, also referred to as lexical chunks or formulaic sequences, are recurrent sequences of words that co-occur more frequently than expected by chance in natural discourse. Various scholars have provided different definitions based on their research focus. Biber et al. (1999) defined lexical bundles as sequences of three or more words that commonly appear together, regardless of their idiomaticity or structural completeness. Wray (2002) described them as sequences stored and retrieved whole from memory rather than generated anew each time, emphasizing their prefabricated nature. Hyland (2008) highlighted their role in shaping meaning and contributing to text coherence by defining them as extended collocations that appear more frequently than expected by chance.

2.2 Research on Lexical Bundles

Research on lexical bundles has a rich history, particularly in the field of applied linguistics, where it has been explored from various perspectives such as defining, identifying, and classifying these multi-word units. With the advent of corpus linguistics, empirical studies flourished, exploring various

13

facets of lexical bundles.

Numerous studies demonstrate that lexical bundle use varies significantly across genres and disciplines. Biber et al. (2004) identified distinct categories of bundles in academic writing (research-oriented, text-oriented, participant-oriented), highlighting their functional roles in creating coherent texts. Further variations are evident in research articles (Hesamoddin, 2017), textbooks (Ribeck & Borin, 2014), and even specialized genres like hotel websites (Fuster-Marquez, 2014), underscoring the context-dependent nature of these units.

The discipline-specific usage of lexical bundles has been documented in studies comparing their prevalence and functions across fields like science, humanities, and social sciences. Cortes (2004) investigated how lexical bundles function differently in biology and applied linguistics, highlighting how the demands of specific academic communities influence lexical choices. Similar studies by Hyland (2008) have underscored the variations in bundle usages between natural and applied sciences, suggesting that these differences are closely tied to the epistemological and communicative practices of the disciplines.

Significant comparative research has investigated how native and non-native speakers differ in their use of lexical bundles. Research suggests that native speakers generally employ a wider and more diverse range of bundles (Chen & Paul, 2010). Non-native writers, however, often exhibit a more limited repertoire, potentially due to factors like L1 influence and limited exposure to authentic language (De Cock, 1998, 2000; Granger, 1998a). This difference is exemplified in studies of Chinese learners, who tend to overuse certain structures like clause-based bundles compared to the phrase-based preferences of native speakers (Pan, 2016; Li & Liu, 2016). Such deviations from native norms can impact the fluency and coherence of non-native writing (Adel & Erman, 2012; Fatih & Hacer, 2016), highlighting the need for targeted instruction and support in this area. These differences underscore challenges non-native speakers face, likely due to limited exposure to diverse lexical bundles and linguistic transfer from their first language.

3. Research Methodology

3.1 Theoretical Framework

3.1.1 Definition of Lexical Bundles in the Present Study

In this study, lexical bundles are defined as recurrent sequences of four words that frequently co-occur in academic texts written by both native and non-native English-speaking sociology students. This focus is justified as four-word bundles offer a balance between specificity and frequency, providing substantial contextual information while remaining frequent enough for robust statistical analysis. Furthermore, four-word lexical bundles encompass a wide range of structural and functional categories, making them an ideal focus for examining how different types of bundles are used by NS and NNS writers. By analyzing these bundles, the study provides a detailed understanding of the structural and functional patterns in academic writing.

14

3.1.2 Classifications of Lexical Bundles

This study is based on two well-established classifications of lexical bundles: Biber's structural classification and Hyland's functional classification. These frameworks provide a comprehensive approach to analyzing the structure and function of lexical bundles within the context of sociology papers written by native and non-native English speakers.

Biber et al.'s (1999) structural classification basically includes noun phrase (NP) based bundles, prepositional phrase (PP) based bundles, and verb phrase (VP) based bundles. Their examples and descriptions are shown in Table 1.

Table 1. Biber et al.'s (1999) Structural Classification

Category	Examples	Description
Noun Phrase-Based Bundles	the end of the, the purpose of the	Often include a noun followed by a prepositional phrase, prevalent for defining and specifying academic concepts.
Prepositional Phrase-Based Bundles	in the context of, at the end of	Provide locational or temporal context, crucial for establishing the scope and framework of academic arguments.
Verb Phrase-Based Bundles	it is important to, is known as a, is expected to	Cover a variety of structures including verb/adjective phrase fragments, passive constructions, and clauses with "that" or "to", instrumental in conveying actions, states, or processes.

Complementing the structural analysis, Hyland's (2008) functional classification provides insight into the roles that lexical bundles play within academic texts. This classification divides lexical bundles into three primary functional categories: research-oriented, text-oriented, and participant-oriented bundles. Their sub-categories, examples and descriptions are shown in Table 2.

Table 2. Hyland's (2008) Functional Classification

Categories	Subcategories	Examples	Description
Research-oriented	Location	in the area of	Indicates spatial or temporal context.

	Procedure	the use of the	Describes methodological steps.
	Quantification	a large number of	Indicates quantities or extents.
	Description	the nature of the	Provides detailed descriptions.
	Topic	the role of	Indicates subject matter.
Text-oriented	Transition signals	on the other hand	Indicates shifts or contrasts.
	Resultative signals	as a result of	Indicates cause-effect relationships.
	Structuring signals	in conclusion	Organizes the text.
	Framing signals	in terms of	Sets the context for arguments.
Participant-oriented	Stance features	it is important to	Expresses the writer's attitudes or evaluations.
	Engagement features	as we can see	Directly involves the reader in the discourse.

3.2 Research Instruments

This study employs two primary research instruments to extract and analyze four-word lexical bundles from academic texts written by sociology students: AntConc and the Log-Likelihood Ratio Test Calculator developed by Xu Jiajin. These tools facilitate a thorough examination of lexical bundle usages, allowing for detailed comparisons between native and non-native English writers.

3.2.1 AntConc

AntConc, developed by Lawrence Anthony, is a versatile corpus analysis tool widely used in linguistic research. The current study utilizes AntConc 3.5.9 to extract and analyze four-word lexical bundles from the MICUSP. This software is freely available and offers a range of functions essential for corpus analysis, including word cluster analysis, frequency analysis, and word collocation analysis (Anthony, 2019).

3.2.2 Log-Likelihood Ratio Test Calculator

To assess the differences in lexical bundle usages between native and non-native English writers, the study employs the Log-Likelihood Ratio Test Calculator developed by Xu Jiajin. This statistical tool is particularly effective for analyzing frequency data and determining whether the observed differences in

16

lexical bundle usages are statistically significant (Dunning, 1994). This tool is preferred for its ability to handle small sample sizes and provide reliable statistical outcomes, essential for determining the significance of observed variations in lexical bundle usage (Dunning, 1994). In this study, the significance level is set at 0.05, meaning that a log-likelihood (LL) value greater than 3.84 indicates a statistically significant difference between the two corpora.

3.3 Data Collection

The data collection for this study is centered on a carefully curated corpus of academic papers from the Michigan Corpus of Upper-Level Student Papers (MICUSP). MICUSP is a prestigious and extensively used corpus in linguistic research, renowned for its extensive compilation of high-quality student writing across various disciplines at the University of Michigan. This corpus is particularly valuable for examining linguistic features such as lexical bundles due to its rich diversity of texts authored by both native (NS) and non-native (NNS) English speakers.

For this study, a total of 40 academic papers were selected from the Sociology category of MICUSP. This selection includes 20 papers written by native English-speaking students and 20 papers by non-native English-speaking students. The papers were randomly extracted to ensure a representative sample, thereby mitigating any potential biases that could influence the findings. This random selection is crucial for ensuring the generalizability and reliability of the study results.

The 20 papers written by native English speakers serve as a benchmark for proficient academic writing in sociology. These papers are anticipated to exhibit a high degree of lexical sophistication and structural complexity, reflecting the linguistic competence and academic proficiency of native speakers. Conversely, the 20 papers authored by non-native English speakers provide valuable insights into how advanced learners of English utilize lexical bundles in their academic writing. These papers are essential for understanding the challenges and strategies employed by non-native speakers in constructing coherent and fluent academic texts.

The data collection process involved the construction of two self-built corpora from the MICUSP Sociology category. This approach ensured a balanced representation of both native and non-native English speakers, allowing for a focused analysis within the discipline of sociology. For the native speaker corpus, the number of texts totaled 20, with 45,620 running words and an average of 2,281 words per text. This corpus contained 5,561 types and 45,950 tokens, resulting in a type/token ratio of 12.10%. Similarly, the non-native speaker corpus comprised 20 texts, 52,182 running words, and an average of 2,609.1 words per text. This corpus included 5,571 types and 52,114 tokens, with a type/token ratio of 10.69%. These parameters highlight the linguistic diversity and complexity within each corpus, providing a robust foundation for analyzing four-word lexical bundles. The data are shown in the table below.

Table 3. Detailed Data of Two Self-built Corpora

Corpus	Number of texts	Number of running words	Average words of texts	Types	Tokens	Type/Token
Native English Sociology Academic Writings	20	45,620	2,281	5,561	45,950	12.10%
Non-Native English Sociology Academic Writings	20	52,182	2,609.1	5,571	52,114	10.69%

The extraction of four-word lexical bundles was conducted using AntConc, a powerful corpus analysis tool (Anthony, 2019). The process involved several critical steps to ensure the accuracy and relevance of the extracted bundles. Initially, the parameters within AntConc were configured to extract four-word sequences that meet specific frequency and dispersion criteria. This setup was essential for identifying meaningful and commonly used lexical bundles. The N-grams function was then activated to identify and list all four-word sequences within the corpus, generating an initial list of potential lexical bundles. Following this, the initial results were filtered to exclude any sequences that did not meet the minimum frequency and dispersion thresholds, focusing the analysis on the most relevant and significant lexical bundles.

After the automated extraction, a manual refinement process was undertaken to ensure the accuracy and contextual relevance of the identified lexical bundles. This process involved eliminating redundancies by consolidating duplicate or overlapping bundles, such as "the results of the" and "the results of this," to avoid inflating the data and maintain the integrity of the analysis. Additionally, lexical bundles that did not convey substantial meaning or were contextually irrelevant were discarded, ensuring that the analysis focused on bundles that genuinely contribute to the academic discourse.

This meticulous approach to data collection and refinement provides a robust foundation for the subsequent analysis of lexical bundle usage in NS and NNS sociology academic papers. By ensuring the accuracy and relevance of the data, this study aims to contribute valuable insights into the linguistic characteristics and differences in academic writing between native and non-native English speakers.

4. Results

4.1 Structural Analysis of 4-word Lexical Bundles in Two Corpora

This section examines the structural characteristics of four-word lexical bundles in the academic writing of native speakers (NS) and non-native speakers (NNS) within the sociology discipline. Following Biber et al.'s (1999) framework, the structural analysis classifies the bundles into three main categories: noun phrases (NP), prepositional phrases (PP), and verb phrases (VP). The distributions of the structural categories of four-word lexical bundles in the NS and NNS corpora are shown in the

figures and table below.

Table 4. Structural Categories of 4-word Lexical Bundles Usages in Two Corpora

Structure	Type (NS)	Percentage (NS)	Type (NNS)	Percentage (NNS)	P Value
NP	366	69.58%	342	64.90%	0.35
PP	97	18.44%	131	24.86%	0.02
VP	63	11.98%	54	10.25%	0.4

4.1.1 Analysis of Noun Phrase Based Bundles in Two Corpora

Noun phrases (NP) constitute the majority of lexical bundles in both NS and NNS corpora, accounting for 69.58% and 64.90% respectively. This prevalence suggests that noun phrases are a crucial component of academic writing in sociology, serving to introduce and elaborate on key concepts and entities. The slight difference in percentage indicates a marginally higher reliance on noun phrases by native speakers. This could be attributed to the native speakers' ability to use more complex noun phrases, enhancing the sophistication and precision of their academic discourse. However, despite the difference in percentages, the p-value of 0.35 indicates that this variation is not statistically significant. This suggests that both NS and NNS writers rely heavily on noun phrases to achieve clarity and precision in their academic discourse.

4.1.2 Analysis of Prepositional Phrase Based Bundles in Two Corpora

Prepositional phrases (PP) represent the second most common category, with NS papers featuring 18.44% and NNS papers 24.86%. The higher usage of prepositional phrases by non-native speakers is statistically significant (p = 0.02). The higher usage of prepositional phrases by non-native speakers may reflect their tendency to rely on simpler syntactic structures. Prepositional phrases are often used to provide contextual information, link ideas, and establish relationships between concepts. The increased usage in NNS writing could also be indicative of a strategy to ensure clarity and coherence, compensating for potential gaps in lexical and syntactic variety.

4.1.3 Analysis of Verb Phrase Based Bundles in Two Corpora

Verb phrases (VP) are the least frequently occurring category in both corpora, comprising 11.98% in NS papers and 10.25% in NNS papers. The p-value of 0.4 indicates that the difference in the usage of verb phrases between NS and NNS writers is not statistically significant. The low frequency of verb phrases aligns with findings from previous studies, which suggest that academic writing, particularly in the social sciences, tends to favor nominalization and other structures over verb-centered expressions. The slightly higher use of verb phrases by native speakers might indicate a greater comfort with employing a variety of syntactic constructions, including those that foreground actions and processes.

4.1.4 Structural Similarities and Differences of Lexical Bundles Usages in Two Corpora

The comparative analysis of the structural categories of four-word lexical bundles in NS and NNS sociology papers reveals both similarities and differences. While both groups predominantly use noun phrases, native speakers exhibit a somewhat higher tendency towards this structure, although this difference is not statistically significant, potentially reflecting their advanced lexical resources and familiarity with academic conventions. Non-native speakers' significantly increased use of prepositional phrases suggests a reliance on structures that enhance the clarity and coherence of their writing. The consistent, albeit minor, presence of verb phrases in both corpora underscores their role in maintaining a dynamic and varied discourse. These findings deepen the understanding of the structural preferences in academic writing across different linguistic backgrounds, emphasizing areas where non-native speakers may diverge from native norms.

4.2 Functional Analysis of the Lexical Bundles in Two Corpora

This section analyzes the functional characteristics of four-word lexical bundles used by native speakers (NS) and non-native speakers (NNS) in sociology academic papers, with a focus on the similarities and differences in their usage. Employing Hyland's (2008) functional classification, the analysis provides insight into how lexical bundles facilitate the construction of academic discourse and reveals differences between NS and NNS. The distributions of the functional categories and sub-categories of four-word lexical bundles in the NS and NNS corpora are shown in the Tables below.

Table 5. Functional Categories of 4-word Lexical Bundles Usages in Two Corpora

Function	Frequency (NS)	Percentage (NS)	Frequency (NNS)	Percentage (NNS)	P Value
Research-oriented	464	88.2%	448	85.0%	0.58
Participant-oriented	32	6.1%	20	3.8%	0.09
Text-oriented	30	5.7%	59	11.2%	0.00

Table 6. Functional Sub-categories of 4-word Lexical Bundles Usages in Two Corpora

Categories	Subcategories	Frequency (NS)	Percentage (NS)	Frequency (NNS)	Percentage (NNS)	P Value
Research-oriented	Location	61	11.6%	49	9.3%	0.25
	Procedure	34	6.5%	16	3.0%	0.01
	Quantification	8	1.5%	20	3.8%	0.02
	Description	44	8.4%	42	8.0%	0.82
	Topic	317	60.3%	321	60.9%	0.89

Text-oriented	Transition signals	21	4.0%	41	7.8%	0.01
	Resultative signals	5	1.0%	8	1.5%	0.40
	Structuring signals	0	0.0%	6	1.1%	0.00
	Framing signals	4	0.8%	4	0.8%	0.99
Participant-oriented	Stance features	20	3.8%	12	2.3%	0.15
	Engagement features	12	2.3%	8	1.5%	0.37

4.2.1 Analysis of Research-oriented Bundles in Two Corpora

Research-oriented bundles form the backbone of academic writing in sociology, facilitating the description and discussion of research activities and findings. Despite similar overall usage rates, the slightly lower percentage in NNS papers (85.0% compared to 88.2% in NS papers) may reflect differences in the integration of research discourse norms. The P-value of 0.58 indicates that this difference is not statistically significant, suggesting that both NS and NNS are similarly reliant on these bundles for structuring research discourse.

Research-oriented bundles are crucial in academic writing for structuring research discourse. The slight difference in overall usage between NS and NNS can be attributed to the more experienced handling of research narratives by native speakers. NNS's slightly lower usage (85.0% vs. 88.2% for NS) could reflect a developmental stage in acquiring the discourse practices of their academic community (Hyland, 2008).

Significant differences were observed in Procedure (NS: 6.5%, NNS: 3.0%, P = 0.01) and Quantification (NS: 1.5%, NNS: 3.8%, P = 0.02). NS's higher use of procedure-related bundles likely reflects their comfort with established research norms and terminologies. In contrast, the higher usage of quantification by NNS may suggest a compensatory mechanism for asserting scientific credibility or a reliance on quantitative evidence to support claims (Hyland, 2008).

4.2.2 Analysis of Participant-oriented Bundles in Two Corpora

Participant-oriented bundles are significantly less frequent in both corpora, with NS using them more extensively (6.1% vs. 3.8% for NNS). These bundles are crucial for expressing personal stances and engaging the reader, suggesting that NS might be more adept at or more comfortable with integrating their voices into academic text, a skill that NNS appear to use with less frequency (Hyland, 2008). The P-value of 0.09, close to the threshold of significance, hints at an emerging pattern where NNS might

benefit from further development in employing these expressive and interactive language tools.

What's more, this comprehensive difference might be related to the rhetorical style of native speakers, who are typically more adept at engaging with their audience through the text, manifesting a sophisticated use of language to position themselves and their arguments (Hyland, 2008). The subtle differences that are not statistically significant in stance features (NS: 3.8%, NNS: 2.3%) and engagement features (NS: 2.3%, NNS: 1.5%) underscore the challenges NNS face in employing language that confidently positions their research perspective and actively engages with the reader.

4.2.3 Analysis of Text-oriented Bundles in Two Corpora

Text-oriented bundles show the most significant difference in usage, with NNS employing them nearly twice as much as NS (11.2% vs. 5.7%). The P-value of 0.00 confirms the statistical significance of this discrepancy. This higher usage by NNS could indicate a greater reliance on explicit textual structuring devices to aid in the organization of their papers and ensure clarity of the discourse. It may also suggest a compensatory strategy by NNS to adhere to expected academic structures and connect ideas coherently, possibly reflecting less familiarity with implicit or cultural aspects of academic rhetoric.

NNS's increased reliance on these bundles, especially in 'Transition signals' (NS: 4.0%, NNS: 7.8%, P = 0.01) and 'Structuring signals' (NS: 0.0%, NNS: 1.1%, P = 0.00), likely reflects a strategic use to enhance textual coherence and logical flow, compensating for less intuitive grasp of complex academic discourse structures (Hyland, 2008).

4.2.4 Functional Similarities and Differences of Lexical Bundle Usages in Two Corpora

The functional analysis illuminates key aspects of how NS and NNS utilize lexical bundles in constructing their academic discourse. While both groups heavily rely on research-oriented bundles to anchor their discussions, differences in the use of participant-oriented and text-oriented bundles reveal deeper undercurrents in their rhetorical approaches and linguistic proficiency. NNS's greater use of text-oriented bundles underscores a strategic emphasis on ensuring textual coherence and clarity, possibly to compensate for less intuitive engagement with the norms of academic discourse in sociology. In contrast, NS's more frequent use of participant-oriented bundles may reflect a higher comfort level with embedding personal stances and engaging directly with readers, indicative of greater rhetorical sophistication or familiarity with academic conventions (Hyland, 2008).

These findings suggest potential areas for pedagogical focus, particularly in helping NNS develop more confidence and skill in employing participant-oriented bundles to express personal stances and engage readers, thereby achieving a more nuanced and persuasive academic style.

5. Conclusion

This study systematically investigates the structural and functional usages of four-word lexical bundles in academic writings by native (NS) and non-native speakers (NNS) in the field of sociology, utilizing a

corpus-based methodology. Through detailed comparative analysis, the research illuminates significant differences and similarities in how lexical bundles are employed by NS and NNS, contributing both to the theoretical understanding of lexical bundles and to practical applications in EAP development.

5.1 Summary of Findings

The findings indicate that NS and NNS generally employ similar types of lexical bundles, particularly research-oriented bundles, which are pivotal in structuring academic discourse. However, notable differences were found in the use of text-oriented and participant-oriented bundles. NNS tend to rely more heavily on text-oriented bundles to enhance textual coherence and logical structure, likely as a compensatory strategy to address potential weaknesses in discourse management (Hyland, 2008). Conversely, NS more frequently utilize participant-oriented bundles, suggesting a higher comfort level with integrating personal voice and engaging with the academic audience.

These disparities emphasize the challenges NNS face, including a narrower range of lexical bundle usages and difficulties in effectively integrating these bundles to achieve fluent and idiomatic academic prose. The structural analysis further revealed that while both groups predominantly use noun phrases, NNS show an increased dependency on prepositional phrases, potentially to ensure clarity and coherence in their texts.

5.2 Implications for Future Studies

This study's findings have practical implications for the teaching design and implementation of EAP, suggesting that enhancing the instruction of text-oriented and participant-oriented bundles could significantly benefit NNS. By fostering a better understanding of these bundles' structural and functional aspects, educators can help NNS develop more comprehensive writing skills, ultimately leading to improved academic capacity.

Future research can integrate qualitative methodologies, such as interviews with academic writers, to provide deeper insights into the cognitive and pedagogical aspects of using lexical bundles. Moreover, future studies can focus on more specific countries, regions, communities or even groups to reduce the influence of too many variables on the results.

In conclusion, this research, by focusing on the use of lexical bundles in sociology papers written by NS and NNS, provides valuable insights for both linguistic theory and pedagogical practice. The findings can inform the teaching design and intervention of EAP. Furthermore, by illustrating how these bundles contribute to the specific rhetorical goals of sociological discourse, this study equips instructors with the knowledge needed to effectively guide NNS students in English academic writing. Ultimately, this research contributes to a more inclusive and effective academic communication environment by equipping NNS writers with the linguistic tools needed to confidently and competently engage in their field.

References

Adel, A., & Erman, B. (2012). *Recurrent Word Combinations in Advanced Learner Writing: Collocations and Formulaic Language*. In M. Götz-Votteler, & S. Herbst (Eds.), *Collocations and*

Other Lexical Combinations in English (pp. 85-106). Narr.

Biber, D. (2006). *University Lectures*. In K. Aijmer, & B. Altenberg (Eds.), *Advances in Corpus Linguistics: Special Issue of International Journal of Corpus Linguistics* (pp. 162-188). John Benjamins.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Longman.

Chen, S. (2023). *A Comparative Study of the Use of Four-Word Lexical Bundles in English Linguistics Doctoral Dissertations by Chinese and Native English Speakers* (Master's Thesis). Jilin University. CNKI. https://link.cnki.net/doi/10.27162/d.cnki.gjlin.2023.004458

Chen, Y., & Baker, P. (2010). *Lexical Bundles in L1 and L2 Academic Writing*. Language Learning, 60(2), 301-337.

Cortes, V. (2004). Lexical Bundles in Freshman Composition. *Applied Linguistics*, *25*(3), 387-411.

De Cock, S. (1998). A Recurrent Word Combination Approach to the Study of Formulae in Learners' Interlanguage. *International Journal of Corpus Linguistics*, *3*(2), 157-188.

De Cock, S. (2000). Repetitive Sequences in Native Speaker and Advanced Learner Spoken and Written Discourse: A Comparative Study. In C. Mair & M. Hundt (Eds.), *Corpus Linguistics and Linguistic Analysis* (pp. 117-133). Rodopi.

Fatih, B., & Hacer, H. (2016). A Comparative Analysis of Lexical Bundles Used by Native and Non-Native Scholars. *Journal of English for Academic Purposes*, *21*, 1-13.

Fuster-Marquez, M. (2014). Four-Word Lexical Bundles and Discourse Functions in a Corpus of Hotel Websites. *English for Specific Purposes*, *33*(1), 45-56.

Granger, S. (1998). Prefabricated Patterns in EFL Writing: Collocations and Formulae. In C. Cowie (Ed.), *Phraseology: Theory, Analysis and Applications* (pp. 145-160). Clarendon Press.

Hesamoddin, F. (2017). Lexical Bundles in Applied Linguistics Research Articles: A Corpus-Based Study of Introduction, Method, and Results Sections. *Journal of English for Academic Purposes*, *29*, 41-53.

Hyland, K. (2008). Academic Clusters: Text Patterning in Published and Postgraduate Writing. *International Journal of Applied Linguistics*, *18*(1), 41-62.

Hyland, K. (2008). As a Result of Reading: Lexical Bundles and Disciplinary Variation. *English for Specific Purposes*, *27*(2), 131-146.

Li, C., & Liu, H. (2016). A Corpus-Based Comparative Study of Lexical Bundles in Applied Linguistics Research Articles by Native and Chinese Expert Writers. *Journal of World Languages*, *3*(1), 41-53.

Li, Z. (2021). A Comparative Study on the Use of Stance Lexical Bundles in Academic Papers by Chinese and English Scholars in the Field of Foreign Language Teaching. *Foreign Language World*, *4*, 40-49.

Lou, Y., & Wang, X. (2021). A Corpus-Based Study of Lexical Bundles in Academic Journals: A

24

Comparative Study of Native and Chinese Writers. *Foreign Language Teaching and Research*, *54*(6), 1010-1022.

Michigan Corpus of Upper-Level Student Papers (MICUSP). (2024). Retrieved from https://micusp.elicorpora.info/

Pan, F. (2016). Structural and Functional Analysis of Lexical Bundles in Mechanics Journal Articles by Chinese and Native English Writers. *Journal of English for Academic Purposes*, *22*, 1-13.

Ribeck, J., & Borin, L. (2014). Identifying Key Word Combinations for Corpus-Driven Analysis of Learner Language: Combining Corpus and Dictionary Methods. *International Journal of Corpus Linguistics*, *19*(4), 443-471.

Shen, T. (2016). An Experimental Research on the Teaching of Business English Oral Expression Based on Lexical Bundles. *Journal of Changchun University of Science and Technology (Social Sciences Edition)*, *29*(4), 146-149.

Shen, T. (2016). The Effects of Lexical Bundle Teaching in Oral Business English Classes. *Journal of Business English*, *7*(2), 45-58.

Si, L. (2020). Lexical Bundle Use and Development in L2 Academic Writing: A Longitudinal Study of Chinese EFL Graduate Students' Theses. *Journal of Second Language Writing*, *49*, 100731.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge University Press.

Xu, X. (2014). *A Corpus-Based Study of Lexical Bundles in Thesis Writing: A Comparative Study of Chinese EFL Learners and English Native Expert Writers of Applied Linguistics* (Master's Thesis). North China Electric Power University.

Anthony, L. (2019). *AntConc (Version 3.5.8) [Computer Software]*. Tokyo, Japan: Waseda University. Retrieved from http://www.laurenceanthony.net/software