

A Corpus-based Analysis of Lexical Chunks in High School Students' English Writing

Shuangrui Yu

China West Normal University, Nanchong, China

Abstract

This study examines the use of lexical chunks in the English writing of 179 senior high school students from seven Sichuan high schools. Using BFSU PowerConc 1.0 software, the research aims to identify the characteristics and shortcomings of lexical chunk usage. Findings reveal that while students frequently use lexical chunks—about one for every five words—the variety of chunks is limited. Polywords are the most commonly used, followed by phrasal constraints and sentence builders, with institutional expressions being the least utilized. Additionally, students often make spelling errors and misuse collocations, and they rarely use institutional expressions effectively. The paper concludes with recommendations for teachers to enhance students' awareness of lexical chunks during second language acquisition, diversify teaching materials, and improve the quality of writing exercises.

Keywords

English Writing; Lexical Chunks; Corpus; High School Students.

1. Introduction

The four language skills—listening, speaking, reading, and writing—are essential in second language acquisition, with writing playing a crucial role in foreign language learning. According to the New English Curriculum Standard for General Senior High School (2017), writing requires students to choose suitable discourse types, vocabulary, and grammatical structures based on their expressive goals, making effective vocabulary use vital. High school English writing is crucial for both the college entrance examination and overall language proficiency, leading many Chinese scholars to explore the challenges related to students' English writing. Key factors affecting writing quality include vocabulary usage and text structure. Students often struggle with vocabulary, frequently memorizing words in isolation and relying on Chinese-to-English translations, which can lead to unnatural language use (Weng, 2011). This problem is compounded by teaching methods that lack contextual vocabulary instruction. This study analyzes the types, frequency, and shortcomings of lexical chunk usage in high school students' English writing, providing recommendations for improvement. The lexical approach, highlighted by Lewis (1993), emphasizes fixed grammatical structures that enhance collocations and coherence in discourse. Understanding lexical chunks is essential for improving students' expression skills and writing quality. Traditional teaching often prioritizes grammar and vocabulary over lexical chunks, especially in writing classes. This research aims to educate students about lexical chunks' role and offer teachers a method to enhance writing abilities.

2. Literature Review

2.1. Brief Introduction of Lexical Chunks

2.1.1. Definitions of Lexical Chunks

At present, many scholars believe that lexical chunks are important and have different definitions of lexical chunks. Becker (1975) proposed that discourse is created by repeating, modifying, and connecting known phrases made up of multiple words, being one of the first to coin the term lexical chunks. He argued that people's daily communication relies on the collocation of previously learned phrases, suggesting that understanding these chunks is fundamental to grasping the language as a whole. After learning lexical chunks, individuals can recall and use memorized phrases, enhancing efficiency in speaking and writing. Pawley and Syder (1983) introduced the concept of "lexicalized sentence stems" and further defined lexical chunks. Research indicates that the average mature English speaker knows hundreds of thousands of these stems. They characterized lexical chunks as language units shorter than sentences with fixed grammar and lexical content, often referred to as terms. Because different lexical chunks have various grammatical and structural constraints, individuals can generate distinct words when organizing the same chunk in their minds.

Although scholars have different definitions of lexical chunks, all of them indicates that lexical chunks are words that are often paired together in real-life contexts of language use.

2.1.2. Classifications of Lexical Chunks

Biber, Conrad, and Cortes (2004) classify lexical chunks into three categories based on their functions. Stance bundles express attitudes and opinions, such as "It is impossible that...". Discourse organizing bundles help introduce or elaborate on topics, often indicating a speaker's intention, like a teacher saying, "Before our class today, I...". They also include discourse markers such as "I mean..." or "You know..." when clarification is needed. Referential bundles describe real entities or details, often used to state and discuss points, like "one of the things..." or "something like that."

Nattinger and DeCarrico (1992) proposed a language teaching approach centered on lexical phrases, dividing them into four categories. Polywords are fixed phrases interchangeable with single words, such as "for instance" or "by the way." Institutional expressions are unchangeable phrases like proverbs, e.g., "No pains, no gains." Phrasal constraints consist of fixed phrases that restrict other parts of the phrase, such as "get out" or "put off." Sentence builders are longer phrases that provide a framework for clearer expression, like "This is not as...as you think." Their classification is objective and will be used as the framework for this study, as polywords and phrasal constraints align with the phrases high school students memorize in English learning.

2.2. Theoretical Foundations

2.2.1. Lexical Approach

Lewis (1993) proposed a theory that differs from traditional teaching methods in his book "The Lexical Approach" in which he said language is not made up of words and grammar alone but is organized through frequently used blocks of language, which are acquired when the language is used. People can learn and use chunks of language correctly to make communication clearer, more fluent and easier. The key point in his book is "chunk," and a "chunk" can be one word or more. He emphasizes the teaching of meaningful chunks, which will help communication, and therefore the study of chunks in student writing can improve the quality of teachers' teaching, based on the theory of lexical approach.

2.2.2. Input Hypothesis

American linguist S.D. Krashen (1985) proposed a well-known and controversial model of second language acquisition, consisting of five hypotheses: the Acquisition-learning Hypothesis, the Natural Order Hypothesis, the Monitor Hypothesis, the Input Hypothesis, and the Affective Filter Hypothesis. Among these, the Input Hypothesis is considered the most crucial. It posits that learners acquire language through comprehensible input that slightly exceeds their current competence level. Krashen introduced the formula “ $i+1$,” where “ i ” represents learners’ present competence, and “ $i+1$ ” is the next level of difficulty. Exposure to language materials at the “ $i+1$ ” level positively influences language development. If learners only encounter material they already know or find too challenging, it hinders acquisition. Krashen also argued that language skills, such as speaking, are acquired naturally through extensive exposure to intelligible language rather than direct teaching. Understanding the theory of language input is essential for effective language acquisition. Therefore, teachers should provide students with authentic, appropriately challenging, and comprehensible language input. Analyzing lexical chunks in students’ compositions offers insights into their actual and effective input.

2.2.3. Output Hypothesis

Krashen emphasizes that in second language acquisition, teachers should go beyond teaching grammar and general activities to include tasks that facilitate natural language acquisition. He asserts that comprehensible input is essential for language acquisition, leading to natural output once students have received enough input. However, research shows that language acquisition is less effective when teachers provide ample input but limited opportunities for output (Swain, 1985). Swain contends that both comprehensible input and output are vital and mutually reinforce each other. In his 1995 work, Swain further outlines the functions of comprehensible output in language learning, including noticing, hypothesis-testing, and metalinguistic reflection. Thus, when high school students produce lexical chunks in their English writing, they consciously focus on the chunks stored in their minds and actively engage in self-testing and revision to enhance their expressions. Output also allows students to evaluate the effectiveness of their input. Therefore, teachers should not only ensure sufficient language input but also create ample opportunities for language output. Consequently, students’ writing can reflect their overall language skills, making the exploration of lexical chunks in their compositions meaningful.

2.3. Previous Relevant Studies on Lexical Chunks at Home and Abroad

Research on lexical chunks is more prevalent abroad. Biber and Barbieri (2007) studied spoken and written university discourse, finding that lexical chunks are frequently used in teaching, particularly in university classrooms, contrasting with earlier studies that focused on writing. They identified three main functions of lexical chunks: stance expression, denotative expression, and discourse organization. Cortes (2007) examined college students’ awareness of lexical chunks in history class assignments, revealing that exposure to chunks in class increased students’ awareness and usage in their writing. Csomay (2013) analyzed the distribution of lexical chunks, noting that stance chunks were used more at the beginning of instruction, referential chunks increased as instruction progressed, while discourse organization chunks were the least utilized.

In China, research has also focused on lexical chunks. Huang and Zhou (2016) highlighted their importance in measuring learners’ language ability and analyzed high school students’ compositions to adjust teaching strategies, improving efficiency. Wang and Zhang (2007), along with Li and Jiang (2023), noted that Chinese college students’ verb chunks lacked depth and were often used colloquially. Liu (2022) explored the lexical approach in English writing teaching, showing that understanding students’ compositions can enhance writing ability.

In summary, while there is considerable research on lexical chunks at home and abroad, more attention has been given to their role in college students' writing and teachers' instruction. There is a need for further exploration of lexical chunks in high school students' compositions and strategies to enhance their awareness.

3. Research Methodology

3.1. Research Questions

This paper aims to get to know high school students' use of lexical chunks in their writing. It can provide some advice for high school teachers' writing teaching by studying the following two main questions:

- (1) What are the characteristics of the use of lexical chunks in high school English writing?
- (2) What are the shortcomings of high school students' use of lexical chunks in English writing?

3.2. Research Corpus

The corpus for this study consists of 179 time-limited essays on the theme of "Family, friends, and people around us" with the title of "The person I admire most" from senior high school students of Nanchong No.1 Middle School, Nanchong High School, Nanchong Baita Middle School, and Yibin No.3 Middle School.

3.3. Research Instruments

BFSU PowerConc 1.0 : It is retrieval software for corpus. With 179 English compositions into the corpus, it can help accurately generate word lists of different lengths and count the number and frequency of certain lexical chunks.

Microsoft Excel: It can form different tables of lexical chunks according to type, function and length, and can form different statistical graphs for analysis as needed.

3.4. Research Procedures

This study consists of three main steps. First, the composition corpus of senior students is collected. Second, the number, frequency, and types of lexical chunks in students' compositions are analyzed. Third, the data is examined to assess the current status of lexical chunk usage in writing. The detailed procedures are as follows:

(1) Corpus selection: A total of 179 compositions written by senior high school students in Nanchong, Sichuan Province, were selected and transcribed into plain text for corpus software retrieval.

(2) Chunk extraction: Lexical chunks are words commonly used together in real language situations. The frequency of chunks in students' compositions is analyzed. Biber and Barbieri (2007) identified a high frequency of 40 occurrences per million words, while this study will extract chunks appearing three times or more, as more occurrences indicate greater value for analysis.

(3) Data analysis: The author used corpus software to extract 2-6 word lexical chunks, which were then imported into Excel for manual screening. The screening process included understandable spelling mistakes. The analysis calculated total word count, total lexical chunks, average words per article, average lexical chunks per article, and frequency of lexical chunk use based on Zhang's (2004) formula. The study follows Nattinger and Decarrico's classification of lexical chunks into four types: polywords, institutional expressions, phrasal constraints, and sentence builders. Statistical data is presented in a percentage pie chart to illustrate the proportion of various lexical chunk types in students' writing.

4. Results and Discussions

4.1. The Overall Distribution of Lexical Chunks in English Writing

In this study, 179 essays from senior students were collected, and the calculation method: (the total number of words - the number of words contained in the lexical chunks) /the total number of lexical chunks is used to show the frequency of lexical chunks. Table 1 presents the types and frequency of lexical chunks usage in the students’ writing samples. The total number of words in the students’ essays is 21,401, so we can know the average number of words per article is 119.56. In terms of the use of lexical chunks, the total number of chunks used is 2,701, and the average number of chunks per essay is 15.09, which means that the average number of chunks in each writing is about 15. According to the formula of frequency of using lexical chunks, the frequency of lexical chunks is 5.41, which means that every 5 words will contain a lexical chunk. From the above data, we can conclude that most students are able to use a certain number of lexical chunks in their writing. J Hou et al. (2018) believed that the use of chunk is not only a feature of first language use, but also of second language use. Through their study, advanced Chinese L2 learners of English can all use certain lexical chunks in their writing, but the difference lies only in proficiency, which is reflected in the chunk coverage.

Table 1. The Overall Distribution of Lexical Chunks in Students’ Writing

Types	Frequency	Proportion	Example
Polywords	1317	48.76%	because of... such as...
Institutional expressions	15	0.55%	No pains, no gains Everything is possible
Phrasal constraint	387	14.33%	give up try to do...
Sentence builder	982	36.36%	Not only...but also It is...that...
Total	2701	100%	

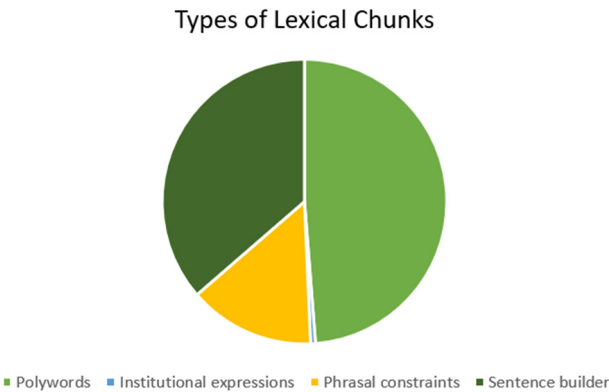


Figure 1. Different Types of Lexical Chunks

There are some reasons why students use different types of lexical chunks. Most of the lexical chunks are fixed phrases that often appear together. After memorizing the lexical chunks, students can quickly extract them when they need to use them.

4.2. Shortcomings in Students' Use of Lexical Chunks in Writing

English writing has always been a weak point not only for students in the process of second language acquisition but also for teachers in the process of teaching English. According to the data analysis, it can be concluded that the main shortcomings of students in the process of using lexical chunks.

First of all, students made some errors in spelling lexical chunks, which indicates that they have not fully grasped some words and lexical chunks. Even if students can retrieve the appropriate chunks from their memory, spelling errors will influence scores and quality of their compositions.

Secondly, students use some collocations incorrectly, which may be influenced by the language structure of their native language. Because of the limited time, students are not able to check out all errors.

Finally, students currently do not develop an awareness of lexical chunks and only write down some common words and phrases according to the teacher's instructions. Students' use of institutional expressions is also very little, indicating that they do not accumulate them carefully in their daily lives. Lexical chunks used by senior students are very simple and have a certain tendency to be colloquial, which affects the quality of their compositions.

5. Conclusion

From the statistics analysis, the frequency of lexical chunks in high school students' English writing is relatively high, one chunk is used for every five words, and the number of chunks is sufficient, with 2701, but the variety of chunks is small. According to the four types of chunks used in the composition, the first one is polywords, which generally have a fixed form and lack flexibility, and the lexical chunks used by students are relatively common. The second is sentence builder, which can provide a certain framework for students' writing, and students seldom make mistakes when using this type of lexical chunk. The third is phrasal constraint, which is easier for students to master and use. Lastly, institutional expressions, which require a lot of accumulation.

What's more, there are also some shortcomings in lexical chunks used by students. First of all, students make some spelling errors in using lexical chunks. Secondly, students use some collocations incorrectly. Furthermore, lexical chunks used by students are very common, and include few types. Students' use of institutional expressions is also very little. It shows that the English writing level of high school students needs further improvement in terms of collocation, word usage and logical thinking.

The study of the use of lexical chunks in English writing texts of high school students shows that teachers need to take some effective teaching measures to improve the quality of students' writing texts. Concerning the limitations mentioned above, the author makes two suggestions for further research. First, the number and genres of samples can be increased, and the range of samples collected should be expanded to better represent high school students' use of lexical chunks in English writing. Second, regarding the subjective factor in the selection process, they need to be further examined and screened by more experienced teachers to reduce the errors in the process of calculating the number of lexical chunks.

References

- [1] Becker, J. (1975). The Phrasal Lexicon. Bolt, Beranek & Newman.
- [2] Biber, D. & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286.
- [3] Huiping Zhang. (2020). On the Features of Lexical Bundles Used by Chinese Beginner Learners of English in Writings. *Foreign Languages Bimonthly*(4):19-25+86+160.
- [4] Jianqin Zhang. (2004). A comparative study of beginners', intermediate learners' and advanced learners' use of lexical phrases. *Foreign Language World*(1):10-14.
- [5] Kaisheng Huang & Xinping Zhou. (2016). Corpus-based Trend Study on Chinese EFL Learners' Lexical Bundle Use. *Foreign Language World*(4):27-34
- [6] Krashen, S. D. (1985). *The Input Hypothesis: Issues and Implication*. Longman.
- [7] Lewis, M. (1993). *The Lexical Approach*. Language Teaching Publications.
- [8] Lifei Wang & Yan Zhang. (2007). A Corpus-based Study of High Frequency verbs in Chinese Learners' EFL Writing. *Foreign Language Teaching and Research*(2):110-116+160-161.
- [9] Nattinger, J.R. & DeCarrico, J.S. (1992). *Lexical Phrases and Language Teaching*. Oxford University Press.
- [10] Swain, M. (1985). *Communicative competence: Some roles of comprehensible input and comprehensible output in its development*. Cambridge University Press.
- [11] Swain, M. (1995). *Three functions of output in second language learning*. Cambridge University Press.
- [12] The Ministry of Education of the People's Republic of China. (2020). *New English Curriculum Standard for General Senior High School: 2017 edition revised in 2020*. People's Education Press.
- [13] Tingting Liu. (2022). Three Dimensions of the Word Block Approach to Teaching English Writing in High Schools. *Journal of the Chinese Society of Education*(12):102.
- [14] Yan Li & Yajun Jiang. (2023). Lexical Bundles in English Academic Discourses by Chinese and Native English Writers: A Corpus-driven Approach. *Technology Enhanced Foreign Language Education* (02), 56-62+113
- [15] Zhixiang Weng. (2011). Analysis of the Current Situation of High School English Writing and Countermeasures - Based on the Perspective of Lexical Articulation. *Fujian Tribune*(S1):240-242.