Asia Pacific Science Press

Journal of Educational Theory and Practice
ISSN(O): 3005-9313
ISSN(P): 3005-9305
https://doi.org/10.62177/jetp.v1i3.85

The Analysis of the Structural and Functional Characteristics of Four-Word Clusters in Experts and Students' Academic Writing in Linguistics

Aoran Zhang

Beijing University of Technology, Beijing, 100124, China

***Corresponding author**:* Aoran Zhang, erinzhang0111@gmail.com

Abstract: This study examines the structural and functional characteristics of four-word clusters in the academic writing of expert and student writers within the field of linguistics. Using a self-compiled corpus and AntConc software, the researcher extracted and analyzed four-word clusters, identifying distinct usage patterns between the two groups. Expert writers showed a strong preference for noun phrase clusters and research-oriented bundles, indicating advanced linguistic competence and a focus on detailed research descriptions. In contrast, student writers demonstrated a more balanced structural distribution and a higher use of text-oriented and participant-oriented bundles, suggesting emerging academic writing proficiency. These findings highlight the need for targeted academic writing instruction, emphasizing noun phrase clusters and research-oriented bundles to guide students towards a more sophisticated writing style. This research provides insights into the teaching of academic English, supporting students' development of more expert-like linguistic practices.

Keywords: Four-Word Clusters; Academic Writing; Corpus Linguistics

Published: Nov 29, 2024

1. Introduction

With the advancement of computer technology and corpus linguistics providing convenience, the study of word clusters has also become a hot topic. One of the important linguistic features of academic texts is the frequent occurrence of word clusters. Some high-frequency word clusters in academic discourse can enhance the naturalness of language expression and highlight the author's identity as an "insider" in a specific linguistic community [1]. Therefore, the study of word clusters in academic English has become a hot topic, and the development of computer technology and corpus linguistics has facilitated the extraction of word clusters by providing ample objective data [2].

Many studies focus on the word clusters that appear in spoken and written language across different fields, treating word clusters as units of meaning in language [2][3][4]. In corpus-based multi-word unit research, word clusters are the starting point of all observations [5]. In fact, the phenomenon of word clusters reflects the prefabricated, conventional, and modular characteristics of language use [6]. This indicates that the use of word clusters is one of the important bases for studying thematic characteristics of texts. Word clusters are integral components of academic texts, possessing unique linguistic features that merit research and exploration. However, most researches have been focused on the comparative study between native and non-native writers, few studies focus on comparing the similarities and differences between expert and student writers.

This study will be based on a self-built corpus of English academic texts, using AntConc 4.2.4 software to systematically examine the similarities and differences in the use of four-word clusters by expert writers and student writers in academic discourse, and to analyze their structural and functional characteristics.

2. Literature Review

Word clusters, also known as "lexical bundles," "word strings," "multi-word sequences," or "multi-word units," are defined as "the most frequent recurring sequences of words in texts" [7][8]. Word clusters are linguistic structures characterized by lexical and grammatical features [9], which can be stored and used as whole units [10], thereby invisibly reducing the burden of language processing and output, making communication faster, more fluent, and more efficient [11].

As for the structural classification of lexical bundles, Biber et al. identified three major structural types: (1) Type 1 bundles incorporate verb phrase fragments. (2) Type 2 bundles incorporate dependent clause fragments in addition to simple verb

phrase fragments. (3) In contrast to Types l and 2, which have clausal components, Type 3 bundles are phrasal [8]. The specific classification of each type is shown in Table 1.

Table 1: The structural classification of lexical bundles

Types	Sub-types	Example Bundles
Type 1: Lexical bundles that incorporate verb phrase fragments	1 a. (connector +) 1st/2nd person pronoun +VP fragment	*you don't have to, I'm not going to, well I don't know*
	1 b. (connector +) 3rd person pronoun +VP fragment	*It's going to be, that's one of the, and this is a*
	1 c. Discourse marker + VP fragment	*I mean you know, you know it was, I mean I don't*
	1 d. Verb phrase (with non-passive verb)	*is going to be, is one of the, have a lot of, take a look at*
	1 e. Verb phrase with passive verb	*is based on the, can be used to, shown in figure N*
	1 f. yes-no question fragments	*are you going to, do you want to, does that make sense*
	1 g. WH-question fragments	*what do you think, how many of you, what does that mean*
Type 2: Lexical bundles that incorporate dependent clause fragments	2 a. 1st/2nd person pronoun + dependent clause fragment	*I want you to, I don't know if ,I don't know why, you might want to*
	2b. WH-clause fragments	*what I want to, what's going to happen, when we get to*
	2c. If-clause fragments	*if you want to, if you have a, if we look at*
	2d. (verb / adjective +) *to*-clause fragment	*to be able to, to come up with, want to do is*
	2e. *That*-clause fragments:	*that there is a, that I want to, that this is a*
Type 3: Lexical bundles that incorporate noun phrase and prepositional phrase fragments	3a. (connector +) Noun phrase with *of*-phrase fragment	*one of the things, the end of the, a little bit of*
	3b. Noun phrase with other post-modifier fragment	*a little bit about, those of you who, the way in which*
	3c. Other noun phrase expressions	*a little bit more, or something like that, and stuff like that*
	3d. Prepositional phrase expressions	*of the things that, at the end of, at the same time*
	3e. Comparative expressions	*as far as the, greater than or equal, as well as the*

As for the functional classification of lexical bundles, there are currently two main classification methods for the functions of lexical bundles. One is three major types proposed by Biber et.al [8], and the other classification is proposed by Hyland [12].

Biber et al. categorized lexical bundles into three major types when studying their usage in university classrooms and textbooks: stance, discourse organizing, and referential bundles [8]. Each type fulfills different functions in specific contexts. Stance bundles provide a frame for the interpretation of the following proposition, conveying two major kinds of meaning: epistemic and attitude/modality. Epistemic stance bundles comment on the knowledge status of the information in the following proposition: certain,uncertain, or probable/possible (e.g.I don't know if; I don't think so). Attitudinal/Modality stance bundles express speaker attitudes towards the actions or events described in the following proposition (e.g. I want you to; I'm not going to). Discourse organizing bundles serve two major functions: topic introduction /focus and topic elaboration/clarification. Referential bundles generally identify an entity or single out some particular attribute of an entity as especially important.

Hyland categorized lexical bundles in academic English discourse into three major types: research-oriented, text-oriented, and participant-oriented bundles [12]. Each major type is further divided into several subcategories. Research-oriented bundles help writers to structure their activities and experiences of the real world. Text-oriented clusters are concerned with the organization of the text and the meaning of its elements as a message or argument. Participant-oriented bundles are focused on the writer or reader of the text. The detailed classification is shown in Table 2.

Table 2: The functional classification of lexical bundles

Broad categories	Subcategories	Examples
Research-oriented bundles	location	*at the beginning of, at the same time, in the present study*
	procedure	*the use of the, the role of the, the purpose of the, the operation of the*
	quantification	*the magnitude of the,a wide range of, one of the most*
	description	*the structure of the, the size of the*
	topic	*in the Hong Kong, the currency board system*
Text-oriented bundles	transition signals	*on the other hand, in addition to the, in contrast to the*
	resultative signals	*as a result of, it was found that, these results suggest that*
	structuring signals	*in the present study, in the next section, as shown in fig.*
	framing signals	*in the case of, with respect to the, on the basis of, in the presence of, with the exception of*
Participant-oriented bundles	stance features	*are likely to be, may be due to, it is possible that*
	engagement features	*it should be noted that, as can be seen*

Research on lexical bundles in academic texts has been deepening both internationally and domestically. Hyland found significant differences in the frequency and preference for lexical bundle usage across different disciplines [13]. For instance, texts in electrical engineering use the most types of lexical bundles, while biology texts use the least. Ma provided a detailed definition of "lexical bundles" and suggested that they can be classified based on structure, function, the number of component words, frequency of occurrence, and tightness [14]. Xu focused on Chinese learners' use of academic lexical bundles in English academic writing and their developmental characteristics [15]. The study found that although Chinese learners use a large number of high-frequency lexical bundles, they share fewer high-frequency bundles with international scholars. Li and Wei explored the discoursal behavior of phrase sequences in academic texts, particularly their roles in stating opinions or facts, reporting, and discourse marking [16]. These sequences not only reflect the textual characteristics of academic communication but also serve as crucial means for propositional and discoursal cohesion.

3. Research Methods

3.1 Research Questions

This study, based on two self-compiled small corpora, examines the structural and functional distribution characteristics of lexical bundles in academic English texts by expert writers and student writers. It aims to address the following two questions:

(1) What are the similarities and differences in the structural types of four-word clusters used by expert writers and student writers?

(2) What are the similarities and differences in the functional types of four-word clusters used by expert writers and student writers?

3.2 The Source of Corpora

This study utilized two sets of corpora, both limited to academic texts on linguistics.

The academic texts written by expert writers were selected from the built-in corpus of the AntCorGen 1.3.0 software, with the subject field specified as Linguistics and the Collocated Fields as Body. All obtained corpora were exported as the source of papers written by expert writers, and ten papers were randomly selected to create a small-scale corpus 1, with a total word count of 29,512 and 28,963 tokens.

The academic texts written by student writers were selected from the BAWE (British Academic Written English Corpus (*ox.ac.uk*)). After downloading the corpus, the subject field was also limited to linguistics, and ten papers were randomly selected to create a small-scale corpus 2. The texts were manually cleaned, removing extraneous information such as authors and references, leaving only the main text, resulting in a total word count of 27,016 and 26,722 tokens.

3.3 Research Procedures

First, extract all four-word clusters from the two sets of corpora using AntConc 4.2.4 software. Four-word clusters are selected because Chen and Baker found that four-word sequences are the most researched length for writing studies [17]. Four-word clusters have a relatively wider range of structures and functions available for analysis; they encompass three-word bundles and are more common than five-word bundles [8][12].

Then, the top 50 four-word clusters were extracted from both the expert writers' corpus and the student writers' corpus, with the least frequent cluster appearing three times. The researcher manually verified the extracted lexical bundles and remove those that are too closely related to the main topic of the papers.

Finally, classify the top 100 lexical bundles extracted from both sets of corpora into structural and functional categories according to the classification methods proposed by Biber et al. for lexical bundle structure and Hyland for lexical bundle function [8][12]. Analyze the differences in frequency and examine the structural and functional characteristics of the four-word clusters.

4. Discussion

This study investigated the structural and functional classifications of lexical bundles in both student and expert data, following the frameworks provided by Biber et al. for structural classification and Hyland for functional classification [8][12]. The results reveal distinct patterns and preferences in the usage of lexical bundles by students and experts, shedding light on their writing practices and linguistic competencies.

4.1 Structural Characteristics of Four-Word Clusters

The structural analysis of four-word clusters revealed notable differences between expert and student writers. According to Figure 1, expert writers predominantly use noun phrase (NP) clusters, comprising 57.9% of their four-word clusters. In contrast, student writers demonstrate a balanced distribution among noun phrase (24%), verb phrase (VP) (30%), and prepositional phrase (PP) clusters (38%). This shows that expert writers use many more noun phrases, which are usually associated with a more specific and sophisticated level of writing in academic works. On the other hand, student writers exhibit a higher frequency of verb and prepositional phrases; hence, their usage pattern is more varied but less specialized.

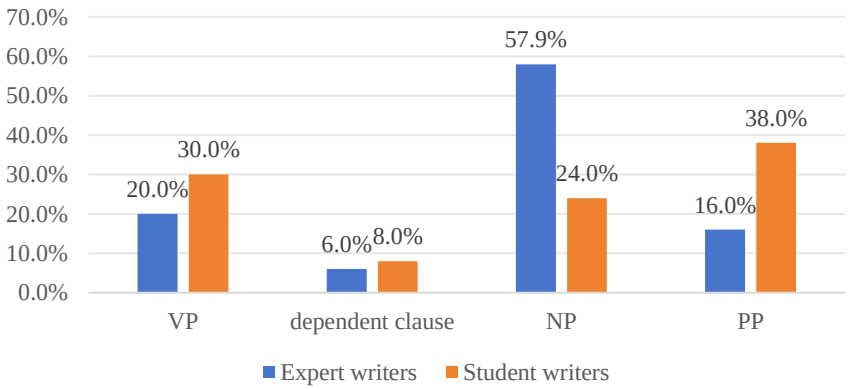

Figure 1: Overall Structural Types of Four-Word Clusters in the Two Corpora

In this respect, the variability reflects student writers apply more varied types of structures than the expert writers, whose work tends to be more specialized. The greater proportional use of verb and prepositional phrases by students represents that they are still dependent on a higher proportion of basic structures and might well be quite challenged by the sophisticated usage of academic writing.

4.2 Functional Characteristics of Four-word Clusters

The functional classification based on Hyland's framework showed that research-oriented bundles and text-oriented bundles are more frequently used both in expert and student writers' work, which generally supports Hyland's viewpoint that research-oriented and text-oriented bundles are the two main functional types in academic discourse [12].

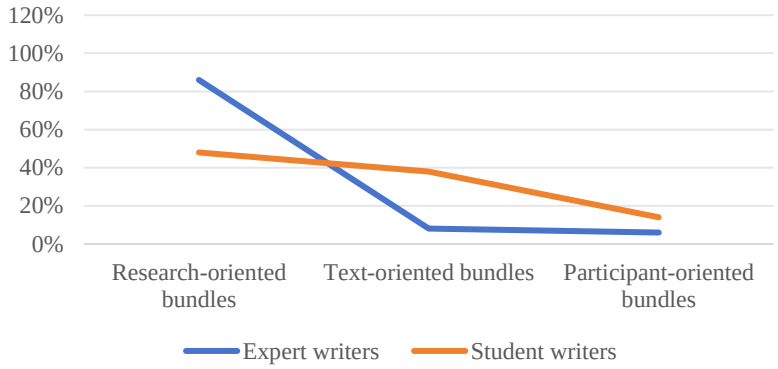

Figure 2: Overall Functional Types of Four-Word Clusters in the Two Corpora

On the other hand, there are also significant differences between the two groups. Expert writers predominantly utilize research-oriented bundles, accounting for 86% of their clusters, as illustrated in Figure 2. This preference underscores their focus on detailing procedures, locations, and quantifications pertinent to their research. Text-oriented bundles are minimally used by experts (8%), and participant-oriented bundles even less so (6%).

In contrast, student writers exhibit a more diversified functional usage. Research-oriented bundles make up 48% of their clusters, showing a less pronounced emphasis on this category compared to experts. Text-oriented bundles are significantly more prevalent in student writing, constituting 38% of their clusters. This higher frequency suggests that student writers are more engaged in structuring their texts and providing discourse signals. Participant-oriented bundles account for only 14% of the clusters in student writing, reflecting a substantially lower rate of pulling the reader or writer into the text than with expert writers.

Table 3: Specific Functional Types of Four-Word Clusters in the Two Corpora

Broad categories	Subcategories	Expert writers		Student writers	
Research-oriented bundles	location	32%	86%	2%	48%
	procedure	30%		16%	
	quantification	8%		0	
	description	16%		18%	
	topic	0		12%	
Text-oriented bundles	transition signals	0	8%	2%	38%
	resultative signals	6%		12%	
	structuring signals	0		8%	
	framing signals	2%		16%	
Participant-oriented bundles	stance features	6%	6%	10%	14%
	engagement features	0		4%	

According to Table 3, it can be found that experts demonstrated a higher frequency and variety of research-oriented clusters. Phrases like “in the context of,” and “the use of the,” indicate their proficiency in discussing research procedures, contextual details, and descriptive elements comprehensively. Text-oriented clusters using words such as “on the other hand,” “as a result of,” and “in the present study” are advanced in showing how professionals structure their text, display results, or present an argument in a well-thought-out manner. These clusters then add to the flow and coherence of the rest of the academic paper. Experts utilize the participant-oriented package to a lesser degree because readers do not participate directly; instead, they are guided by the organized arguments and analyses in each subsection.

While students utilized research-oriented clusters to structure their academic discourse. Phrases like “in the context of,” “the use of the,” and “the role of input” reflect their efforts to describe locations, procedures, and topics relevant to their research. Text-oriented clusters such as “on the other hand,” “as a result of,” and “in the present study” reveal that students seek to structure their text coherently, guiding transitions as well as expounding on the results. These clusters are crucial for maintaining coherence and guiding readers through their arguments. The use of participant-oriented clusters like “it should be noted that” points to students’ keenness to relate or locate what they are saying within the context of an audience and, at the same time, express their stance. These clusters are essential to connect with readers and stress critical points.

5. Conclusions

The comparison of four-word clusters in expert and student academic writing indicates a difference about structural as well as

functional characteristics. Expert writers prefer noun phrase clusters and research-focused bundles, indicating their higher degree of linguistic competence and attention toward rich detailing in research information. In contrast, student writers show a much less biased distribution of the frequency of structural types and higher exploitation of text-oriented and participant-oriented bundles, which indicates an increasing competency in the use of bundles towards academic writing.

These findings have significant implications for academic writing instruction. Educators should emphasize the importance of noun phrase clusters and research-oriented bundles to students, guiding them towards more sophisticated and focused writing practices. Additionally, students should be encouraged to refine their use of verb and prepositional phrases, aligning more closely with the practices of expert writers.

Future research might extend these studies by examining the relationship between such changes and disciplinary variation and by analyzing the effectiveness of interventions designed to facilitate the development of such structures in student writing. By understanding and addressing the distinctive characteristics in lexical bundle usage between expert and student writing, a clearer picture is conceivable of how to support students on their journey towards academic writing.

References

[1] Wang, F.R.& Wang, H.L.(2015).A corpus-based comparative study of academic lexical bundles in linguistics and engineering research articles.*Foreign Language World(02)*,16-24.

[2] Jiang X.Y.(2018).A study of the structural characteristics of the 4-word clusters in Chinese scholars' SCI academic papers. *Foreign Language Learning Theory and Practice(02)*,24-30.

[3] Sinclair, J. (1996). The search for units of meaning. *TEXTUS: English Studies in Italy*, 9(1).

[4] Wei, N.X.(2007).Phraseological characteristics of Chinese learners' spoken English: Evidence of lexical chunks from COLSEC. *Modern Foreign Languages(03)*, 280-291+329-330.

[5] Yang, Y.Y(2013). A corpus-based study on four-word clusters in English for science and technology papers. *Foreign Language Learning Theory and Practice(04)*, 45-52+95.

[6] Li, W.Z.(2007).The Word Clusters in China English News Article Corpus.*Foreign Languages in China(03)*,38-43.

[7] Biber, D., S. Johansson, G. Leech & E. Finegan. (1999). *Longman Grammar of Spoken and Written English*. London: Pearson Education Limited.

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

[9] Yang, H.Z.(2002).*An Introduction to Corpus Linguistics*. Shanghai: Shanghai Foreign Language and Education Press.

[10] Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied linguistics*, 21(4), 463-489.

[11] Ma, G.H.(2009).Lexical bundles in L2 timed writing of English majors.*Foreign Language Teaching and Research(01)*,54-60+81.

[12] Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International journal of applied linguistics*, 18(1), 41-62.

[13] Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for specific purposes*, 27(1), 4-21.

[14] Ma, G.H.(2011).Lexical chunks: Definition, categorization and identification. *Journal of PLA University of Foreign Languages(01)*,1-4+127.

[15] Xu, F.(2012).The Use and Developmental Features of Lexical Bundles in Chinese Learners' English Academic Writing.*Foreign Languages in China(04)*,51-56.

[16] Li, J.J. & Wei, N.X.(2013).A study of discourse act sequences in academic texts. *Foreign Language Teaching and Research(02)*,200-213+318-319.

[17] Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing.