

A Corpus-based Study of Lexical Bundles in Oral Presentations of Thai EFL Undergraduate Students

SORANUT KUMDEE
KANOKPOJ KHAKHAI*

Department of General Education, Faculty of Sciences and Health Technology, Navamindradhiraj University, Thailand

***Corresponding author email: kanokpoj.k@nmu.ac.th**

Article information	Abstract
Article history: Received: 20 Oct 2025 Last revised: 9 Jul 2026 Accepted: 18 Aug 2026 Available online: 24 Aug 2026	Lexical bundles play an important role in language production, improving individuals' language proficiency (Hyland, 2008). While most previous studies have explored lexical bundles used in written discourse, little attention has been paid to analyzing lexical bundles employed in oral presentations. The present study aimed to investigate the use of lexical bundles in 30 oral presentations delivered by Thai English as a Foreign Language (EFL) undergraduate students. Following Biber et al.'s (2004) taxonomy, the findings indicate that four-word lexical bundles serve various discourse functions. Discourse organizers were found to be the most dominant functional category, primarily used to structure ideas, manage the flow of information, and maintain coherence in spoken academic discourse. It was also found that verb phrase fragments were the most frequent structural pattern, reflecting students' preference for formulaic expressions that support fluency and speech planning. However, the absence of several functional and structural categories suggests a limited spoken linguistic repertoire of academic formulaic language. Therefore, explicit instruction and practice in using a wider range of lexical bundles are recommended to enhance students' fluency, coherence, and academic expression in spoken discourse.
Keywords: Lexical bundles Oral presentations Thai EFL undergraduate students Spoken discourse Corpus linguistics	

INTRODUCTION

Oral presentations in English for Academic Purposes (EAP) courses in the higher education context play an important role in English language learning among undergraduate students (Bankowski, 2010). In particular, preparing for in-class presentations can help English as a Foreign Language (EFL) students improve their English skills (Brooks & Wilson, 2014). The commercial materials chosen (e.g., presentation textbooks) also play a role in the impact oral presentations have. Some guidebooks claim to promote effective or perfect presentations, but the link to contextual factors is not clear (Zareva, 2024). Several presentation guidebooks provide a variety of sensible suggestions for presentation improvement, including design, rehearsal, use of voice, body language, idea connection, questions and answers, and even the use of the equipment during the presentation (e.g., Bowden, 2023; Farrington, 2022; Lowe & Pile, 2006).

In commercial textbooks, language expressions are provided in each of the chapters, which are categorized according to functional aspects (e.g., opening and closing, introducing the topics, giving numerical information, and making a conclusion). They tend to include word, phrasal, and sentential elements for generic use. Nevertheless, how could a teacher know whether their students have acquired these phrases from a textbook? What are the typical language expressions employed by students in group presentations? These questions act as a catalyst for more investigation into the way students use language expressions to present their topics. In the context of higher education, oral presentations are typically delivered in academic settings and vary across disciplines (Smakman, 2024). Therefore, further research is warranted into linguistic features and discursal elements associated with oral presentations in both English for General Academic Purposes (EGAP) and English for Specific Academic Purposes (ESAP) contexts. With respect to EFL students, academic oral presentations could be arduous tasks about which students are greatly concerned due to the diffidence in their English language skills (e.g., Kurakan, 2021; Naser & Isa, 2021). Current knowledge suggests that the improvement of oral presentation skills should center on the language needed to give the presentation as opposed to prioritizing other topics.

With the advent of corpora, defined as large, systematically compiled electronic collections of the natural occurrences of spoken and written texts for a variety of linguistic analyses, a great number of research studies have given profound analysis of language in both generic and specific discourses. Increasingly, lexical bundles (LBs), one of the phrasal units investigated, have been given scholarly attention. LB studies primarily investigate how diverse users (such as L1, ESL, or EFL students) naturally use recurring four-word sequences across different registers (Hyland, 2008). Examining the natural use of language either in oral or written communication could shed light on how the students use the language within the same discourse community, (i.e., in-class academic oral presentations). The results from LB research can be utilized as key information for language improvements. Besides, Schmitt and Carter (2004) posited that effective communication can be enhanced by LBs. Also, pedagogical implications from LB research can be applied to improve students' communication and fluency (e.g., Biber & Barbieri, 2007; Biber et al., 2004; Wray, 2013).

Nevertheless, research on LBs has been more prevalent in written discourse than in spoken one. Beginning in the late twentieth century, initial studies primarily examined academic writing such as research articles and textbooks, focusing on aspects such as students' fields of study, comparisons of language features between L1 and L2 users, and language registers (e.g., Biber et al., 1999, 2004; Cortes, 2004). Still, there have been few LB studies into spoken registers (e.g., Biber & Barbieri, 2007; Panthong & Poonpon, 2020). As oral presentations are ubiquitous in tertiary education in Thailand, it is worth examining multi-word sequences produced by Thai EFL undergraduates during the course of those presentations. Regarding the methodological gap, our study seeks to enrich the LB literature with detailed information about LB use in oral presentations in the Thai university context. This leads to the primary objective of the present study, which is to examine the four-word LBs together with their functions and structures within our specialized corpus of oral presentations. There are two research questions as follows:

- (1) What is the functional distribution of four-word lexical bundles in oral presentations delivered by Thai EFL undergraduate students?
- (2) What is the structural distribution of these four-word lexical bundles in oral presentations delivered by Thai EFL undergraduate students?

LITERATURE REVIEW

Prevalent in both functional and structural analyses of language features in different contexts, LBs have been known by a range of terminologies, such as lexical phrases, formulas, routines, fixed expressions, prefabricated patterns, and N-grams (Biber & Barbieri, 2007; Cortes, 2015). The term lexical bundles was first introduced in the work of Biber et al. (1999), *The Longman grammar of spoken and written English*, which stresses the importance of multi-word sequences existing in academic conversation as well as academic writing. The key definition of LBs refers to a series of words frequently co-occurring in a number of utterances or sentences forming a specific register in both written and spoken communication (Biber & Barbieri, 2007). These word sequences sometimes range from three to six words and often serve as building blocks of discourse (e.g., Biber et al., 1999; Biber et al., 2004; Laosirattanachai & Laosirattanachai, 2024).

Early investigations into LBs provided different criteria for identifying multi-word sequences. Thus, certain studies considered these multi-word sequences to be idiomatic expressions. However, it was Biber et al. (1999) who argued that “lexical bundles are recurrent expressions, regardless of their idiomaticity, and regardless of their structural status.” (p. 990). To date, recent works on multi-word sequences (e.g., Barbieri, 2018; Panthong & Poonpon, 2020; Wang & Csomay, 2024) regarded them as non-idiomatic expressions, parallel to the definition given by Biber and Barbieri (2007), which viewed them as incomplete structures with non-idiomatic properties but functioning as discourse construction in spoken-written communication. Consequently, our current study also follows the definition from Biber and Barbieri (2007). This is because idiomatic expressions may occur unnaturally in communication since they must be produced on purpose, as opposed to natural discourse production and pragmatic functions (Biber et al., 1999).

To our best knowledge, interlocutors tend to select an idiom from their existing linguistic repertoire to engage in a conversation often to place emphasis on a particular point, express evaluation, or convey an emotional stance. Idioms are therefore used in association with speakers’ verbal, communicative intentions, which appear to be meaning-driven. For example, the idiom *kick the bucket* is specific to the context involving death; however, it sounds odd unless it is used in such contexts. Unlike LBs, which emerge naturally and recur frequently in discourse, idioms are typically selected based on their meaning and their suitability for a particular context.

Enhancing production of clausal elements in both spoken and written communication, LBs are classified into different discourse functions for distinct categories. Originally, the taxonomy of functions was developed by Biber et al. (1999). This taxonomy did not quite capture other

crucial sub-functions occurring within a register. Later, Biber et al. (2004) put forward a more comprehensive taxonomy of LBs commonly found in both spoken and written texts (e.g., in-class instruction, academic conversation, and textbooks). This taxonomy has gained popularity and has since been applied to a number of studies on spoken discourse (e.g., Biber & Barbieri, 2007; Kashiha & Heng 2013; Panthong & Poonpon, 2020). By and large, this more sophisticated functional taxonomy of spoken and written academic LBs (Biber et al., 2004) is divided into four functions: (1) stance expressions, (2) discourse organizers or discourse organizing functions, (3) referential expressions, and (4) special conversational functions, as seen in Table 1 below.

Table 1
Functional types of lexical bundles (Biber et al., 2004)

Functions	Sub-functions	Examples
1. Stance expressions	1.1 Epistemic stance	<i>the fact that the I don't know why</i>
	1.2 Attitudinal/modality stance	
	a. Desire	<i>I'd like to know do you want to</i>
	b. Obligation/directive	<i>I want you to it is important to</i>
	c. Intention/prediction	<i>we are going to it's going to be</i>
2. Discourse organizers	d. Ability	<i>I'm able to to be able to</i>
	2.1 Topic introduction	<i>I'd like to talk going to talk about</i>
	2.2 Topic elaboration/clarification	<i>so that we can has to do with</i>
3. Referential expressions	3.1 Identification/focus	<i>that it is a one of the most</i>
	3.2 Imprecision	<i>or something like that it looks like a</i>
	3.3 Specification of attributes	
	a. Quantity specification	<i>a lot of things so many of them</i>
	b. Tangible attributes	<i>in the form of be the part of</i>
	c. Intangible attributes	<i>a kind of the a great deal of</i>
	3.4 Time/place/text reference	
	a. Place reference	<i>in some areas of of the United States</i>
	b. Time reference	<i>at the same time at the end of</i>
	c. Text deixis	<i>as shown in the shown in this figure</i>
	d. Multi-functional reference	<i>the middle of the the end of the</i>

Functions	Sub-functions	Examples
4. Special conversation functions	4.1 Politeness	<i>thank you very much</i> <i>thank you for listening</i>
	4.2 Simple inquiry	<i>why do you think</i> <i>what are you doing</i>
	4.3 Reporting	<i>I said to him</i> <i>he told me that</i>

The first function, stance expressions or stance markers/bundles, refers to those used to express speakers' opinions and knowledge, along with their attributes. Additionally, there are two sub-functions that fall within this first function. The first sub-function, an epistemic stance, is generally used to convey individuals' viewpoints, such as certainty, proposition, judgement, and attitudes. The second sub-type, an attitudinal/modality stance, covers speakers' ways of expressing desires, obligations, directions, intentions and predictions.

The second function, discourse organizers, refers to sequences of words used to improve coherence between ideas or information through discourse so that listeners or readers can easily follow what speakers or writers want to present. This helps mark differences between ideas or information within the text in an explicit way. Like epistemic stances, discourse organizers are also divided into two sub-functions: topic introduction and topic elaboration or clarification. Words falling into topic introduction are used to guide readers or writers to different parts of both spoken and written texts. Like words or phrases for transition signals, they help introduce and shape the structures of the texts. In addition to topic introduction, there are many bundles used for elaborating or clarifying the topic. These bundles are used to provide readers or writers with more details or information on the topics that have already been mentioned.

The third function, referential bundles, is used to elaborate the details of things (e.g., physical items, abstract concepts, place, and time) which are associated with the context of the discourse. This third type of LBs is categorized into four sub-types: identification or focus, imprecision, specification of attributes, and time, place, or text reference. Identification or focus is used to emphasize key ideas or information following its initial mention, often with additional details or clarification. These bundles are generally used to highlight and summarize the importance of the topic. Imprecision, in this way, is used to show vagueness of details speakers or writers give. Specifying attributes are bundles used to emphasize the importance and characteristics of details or information given in the discourse. This sub-type can be divided into three aspects: quantity specification, tangible attributes, and intangible attributes. The last sub-type of referential bundle is time, place, or text reference. These subtypes are used to address temporal and spatial modifications as well as the location of the text itself, including visual imagery.

Another function is special conversational function. It encompasses a range of pragmatic aspects pertinent to communication. There are three sub-types as follows: politeness, simple inquiry, and reporting. Politeness refers to bundles representing polite requests, thanking, and apologies. Simple inquiry covers close-ended and open-ended questions whereas reporting, in this way, means clausal structures or verbs phrases like reporting verbs that are used to point to any interlocutors in a discourse context.

Structurally, LBs can be categorized into three grammatical functions (Biber et al., 2004), as illustrated in Table 2 below. The first function is known as verb-phrase fragment, of which there are seven variations of sub-functions as follows: 1st/2nd person pronoun + VP fragment, 3rd person pronoun + VP fragment, discourse marker + VP fragment, VP (with non-passive verb), VP (with passive verb), yes-no question fragments, and WH- question fragments. The second grammatical function is dependent clause fragments, which can be divided into the following five sub-categories: 1st/2nd person pronouns + dependent clause fragments, WH- clause fragments, If-clause fragments, to-infinitival fragments, and that-clause fragments. The last grammatical function is the prepositional phrase fragment. Its sub-functions are as follows: noun phrases with of-phrase fragments, noun phrases with other post-modifier fragments, other noun phrase expressions, prepositional phrase expressions, and comparative expressions.

Table 2
Structural types of lexical bundles (Biber et al., 2004)

Types	Structures	Examples
1. Verb phrase fragments	1.1 (Connector) + 1st/2nd personal pronoun + VP fragment	<i>I'd like to know</i> <i>I'm not going to</i>
	1.2 (Connector) + 3rd personal pronoun + VP fragment	<i>it's going to be</i> <i>and this is a</i>
	1.3 Discourse marker + VP fragment	<i>you know it was</i> <i>I mean I don't</i>
	1.4 Verb phrase (with non-passive verb)	<i>has to do with</i> <i>be the part of</i>
	1.5 Verb phrase with passive verb	<i>shown in this figure</i> <i>can be used to</i>
	1.6 Yes-no question fragments	<i>do you want to</i> <i>are you going to</i>
	1.7 WH-question fragments	<i>what do you think</i> <i>what are you doing</i>
2. Dependent clause fragments	2.1 1st/2nd personal pronoun + dependent clause fragment	<i>I want you to</i> <i>I don't know why</i>
	2.2 WH-clause fragments	<i>what I want to</i> <i>what's going to happen</i>
	2.3 If-clause fragments	<i>if you want to</i> <i>if we look at</i>
	2.4 (Verb/adjective) to-clause fragment	<i>to be able to</i> <i>to come up with</i>
	2.5 That-clause fragments	<i>that I want to</i> <i>that there is a</i>
3. Noun phrase and prepositional phrase fragments	3.1 (Connector) noun phrase with of-phrase fragment	<i>one of the things</i> <i>a little bit of</i>
	3.2 Noun phrase with other post-modifier fragment	<i>a little bit about</i> <i>those of you who</i>
	3.3 Other noun phrase expressions	<i>or something like that</i> <i>a little bit about</i>
	3.4 Prepositional phrase expressions	<i>at the end of</i> <i>at the same time</i>
	3.5 Comparative expressions	<i>as far as the</i> <i>is greater than the</i>

Regarding the contributions of LB research, according to Biber and Barbieri (2007), structures of any bundles are incomplete and do not have any figurative meaning; nonetheless, they usually yield key functions of discourse in both spoken and written communication. With this, it is established that LBs have an influence on whether a person effectively communicates (Schmitt & Carter, 2004). In view of language as natural, formulaic elements (Sinclair, 2004), LBs were regarded as important for language learning since they are prevalent in everyday contexts rather than specific ones. In some academic spheres (i.e., second language acquisition and psycholinguistics), LBs are believed to instill multi-word sequences in a person's long-term memory (Ellis et al., 2008; Wray & Perkin, 2000). Such contributions of LBs related to long-term memory could lead to the application of LBs to support language pedagogy. Another LB contribution is that they are able to connect the two discourses between speakers and listeners (Wray & Perkin, 2000), which potentially enhance language fluency (Hyland, 2008). It is also evident that the use of LBs can reflect the language abilities of a language user (Panthong & Poonpon, 2020). These contributions have renewed interest in LB research.

Currently, the extensive body of literature on LBs tends to focus more on written genres as opposed to spoken discourse (Laosrirattanachai & Laosrirattanachai, 2024). A majority of studies have focused on academic writing, such as research and academic textbooks (e.g., Biber et al., 2004; Farhang-Ju et al., 2024; Gong et al., 2025; Hyland, 2008; Li et al., 2020; Öztürk & Taşçı, 2023; Wright, 2019). A plethora of LB studies in written communication may stem from the fact that written genres have a strong relationship to academic discourse compared to spoken communication (Panthong & Poonpon, 2020). However, we believe that the focus should gradually shift to spoken communication because LBs have potential to improve communication fluency (Hyland, 2008) and effective communication (Schmitt & Carter, 2004). Thus, there should be more investigations into LBs in spoken genres in diverse contexts since spoken LB analyses have been understudied (e.g., Heng et al., 2014; Laosrirattanachai & Laosrirattanachai, 2024; Panthong & Poonpon, 2020).

To shed new light on the actual use of LBs, investigating university students' LBs in spoken academic discourse can provide valuable insights into how these LBs are structured and applied in this context. Still, research into oral presentations of Thai EFL students in various contexts has often focused on identifying speaking anxiety as a primary challenge (e.g., Chinpakdee, 2015; Kalra & Siribud, 2020; Kurakan, 2021). To better understand the functional and structural patterns of language use, our research on LBs in oral presentations can help identify specific linguistic elements that contribute to syntactic and discourse-related challenges.

Regarding previous LB studies highlighting spoken discourse, Heng et al. (2014) analyzed LBs in group discussions in a university context. The results revealed a wide range of structural and functional LBs used in group discussions. Pedagogical implications suggested including LBs, verb-phrases, noun-phrases, and prepositional phrase structures in language lessons for supporting in-class discussion skills. Shin and Won (2024) provided a comprehensive study of LB comparisons between L1 and L2 learners in the context of argumentation of both written and spoken modes. Their results showed that Korean university students often used nominal

and prepositional sequences in the written mode whereas clausal structures were used in the spoken mode. The results also pointed out a large number of attitudinal or modal expressions used, along with some common errors produced during the two communications. However, in the Thai EFL context, limited attention has been given to spoken LBs in academic oral presentations, particularly regarding their functional and structural patterns. Therefore, this study aims to broaden the understanding of how Thai EFL university students use LBs in oral presentations. This is done by examining the extent to which these functions and structures are applied, and by identifying grammatical and discourse-related issues to better understand their use of formulaic language.

METHODS

Research design

In corpus linguistics, two main approaches that are commonly employed are corpus-based and corpus-driven (Tognini-Bonelli, 2001). The former is deductive, using corpus data to test pre-established hypotheses or arguments. In contrast, the latter is inductive, drawing insights directly from the corpus to explore under-researched areas and uncover patterns beyond pre-determined assumptions. To achieve the research objectives, a corpus-based approach was employed to investigate LBs used by Thai EFL learners in oral presentations because the analysis of LBs requires a pre-determined set of categories and criteria derived from previous research, rather than ones that emerge during the analysis. Therefore, the present study adopted a corpus-based approach because it is well-suited for an analysis of the structural and functional characteristics of LBs in the compiled corpus. This decision is supported by previous research demonstrating the effectiveness of the corpus-based approach in analyzing LBs in both spoken and written discourse (e.g., Öztürk & Taşçı, 2023; Siricharoen & Wijitsopon, 2020).

Participants, individual presentations and data collection

We employed a purposive sampling technique to recruit 30 participants, all of whom were first-year students enrolled in an undergraduate English course in the academic year 2023–2024 at a university in Bangkok, Thailand. None of the participants had studied in an English as a medium of instruction (EMI) program or pursued formal English education in an English-speaking country. Their English proficiency was assessed at CEFR level B2 (independent user) based on standardized tests, including TOEFL, IELTS, and the Oxford Online Placement Test.

In this undergraduate English course, an oral presentation is a compulsory component of the curricula. This task is worth 10 marks. Each participant was assigned an individual presentation lasting approximately 5–7 minutes on topics related to scientific breakthroughs in health sciences, general sciences, or psychology. The suggested oral presentation structure (i.e., introduction, body, and conclusion) had been taught before students delivered their presentations. Language expressions in presentations were expected to be articulated given that oral presentations require a range of discourse functions.

Prior to recording, the research objectives were explained to the participants, and informed consent was obtained. The study followed ethical guidelines and received approval from the university's Institutional Review Board (IRB) under COA No. 129/2566. Participant anonymity was strictly protected throughout the research. They were assured of confidentiality and that their responses would be used solely for research purposes. This project was carried out in accordance with the World Medical Association Declaration of Helsinki as well as the Belmont Report.

Oral presentations were recorded in the MP3 files using a digital voice recorder and subsequently transcribed into the TXT files for the corpus-based analysis. This approach allowed for a systematic examination of LBs employed in students' spoken academic discourse.

Data transcription

We employed Transkriptor, an online program, to transcribe the recorded oral presentations from audio to text. This AI-based program is deemed appropriate for transcribing non-native speech, as evidenced by Limgomolvilas (2024), who successfully used it to convert Thai EFL oral presentations into written texts. Despite its effectiveness, careful attention was required when reviewing the transcriptions due to potential phonetic and grammatical issues generated by the program.

To address phonetic issues, all recordings were carefully listened to, as the program could misinterpret words due to mispronunciations or strong accents, resulting in incorrect substitutions. For example, the word "nice" was occasionally misinterpreted despite being clearly articulated by a presenter. Additionally, as an AI-based program, the program sometimes automatically could correct grammatical errors. To preserve the authenticity of the participants' language, all transcriptions were manually reviewed and refined, ensuring that the final scripts accurately reflected the original spoken language.

Finally, the refined 30 DOCX files were converted into the TXT format for analysis using AntConc 4.3.1 (Anthony, 2024) a concordance computer-based program, facilitating a corpus-based investigation of LBs used in the participants' oral presentations.

Corpus compilation

The students' individual presentation corpus (hereafter SIPC) was compiled. SIPC consists of 30 individual oral presentations, each lasting approximately 5–7 minutes. The Q&A sessions following the presentations were excluded to ensure that the corpus fully represented students' LBs in spoken academic discourse. The corpus contains 20,923 tokens. Despite the relatively small size, it can represent classroom discourse, reflecting the articulation of LBs in oral presentations in the particular context. The findings could also offer pedagogical implications used for enhancing students' oral linguistic repertoire in academic presentations in this particular context. For ease of data management and analysis, the 30 presentations were coded as OR1–OR30.

Data analysis

Dan et al. (2024) reviewed 28 previous studies on LBs and found that most research examined bundles from structural and functional perspectives. The current study analyzed LBs structurally and functionally not only to identify their patterns in students' oral presentations but also to understand how these bundles contribute to organizing discourse, signaling stance, and facilitating interaction in spoken academic English.

When it comes to identifying LBs in a corpus, researchers tend to employ criteria deemed appropriate for their specific research objectives. Regardless of the type of corpus, corpus linguists typically consider three key aspects: cut-off frequency, distributional range, and LB length (e.g., Narkprom & Phoocharoensil, 2022; Siricharoen & Wijitsopon, 2020). These criteria help ensure that the identified bundles are both statistically significant and representative of the patterns in the corpus.

Next, determining the distributional range in which LBs occur is another important criterion for deciding whether to include them in a list. Previous studies have shown that most research employing large corpora considered both a normalized rate of occurrence and the number of texts in which a LB appears (e.g., Biber et al., 1999; Biber et al., 2004; Biber & Barbieri, 2007; Narkprom & Phoocharoensil, 2022). Typically, these studies applied a normalized occurrence rate of 10–100 times per million words, combined with the requirement that the bundle occurs in 5–20 different texts.

However, in studies with smaller and specialized corpora (e.g., Siricharoen & Wijitsopon, 2020), a lower threshold of 1% of the total texts was used to account for the limited size of the dataset. Given that the determination of LBs can be arbitrary and that our corpus contains only 30 transcriptions, we proposed a minimum frequency cut-off rate is at least three LBs out of the entire LBs identified in the current corpus, as the criterion for inclusion in the analysis. In other words, the frequency cut-off rate was set at three, meaning that each LB had to be orally produced by at least three students.

In terms of LB length, we focused on four-word LBs for the following reasons. Regarded as multiword sequences, three-word LBs are normally part of four-word LBs and four-word LBs can provide more meaningful insights into structures and functions (Cortes, 2004).

The taxonomy of LBs proposed by Biber et al. (2004) was employed in the current study for the following reasons. Firstly, this taxonomy is based on analyses of large corpora of conversation and academic prose, which are considered appropriate for examining the 30 oral presentations collected in the current study. This is due to the nature of academic presentations, which combine elements of conversation and academic prose and are typically categorized as an undergraduate genre. As Hyland (2009) argued, oral presentations are similar to lectures in that they aim to convey information and knowledge in a relatively informal, monologic, and conventional style. Secondly, this taxonomy provides a comprehensive structural and functional classification of LBs in oral presentations. English for oral presentations is considered a specialized discourse that incorporates features of both academic and conversational

speech. Biber et al.'s (2004) taxonomy classifies LBs from both functionally and structurally, making it suitable for an analysis of language use in academic oral discourse. Thirdly, Biber et al.'s (2004) taxonomy has been employed in previous studies on spoken LBs, such as speech (Flinn, 2023) and academic lectures (Kashiha & Heng, 2013).

The concordance software employed for identifying LBs according to the aforementioned procedural criteria set was AntConc version 4.3.1 (Anthony, 2024), the current version used for data analysis. This software allows the extraction of LBs from the compiled corpus, SIPC, using the N-gram function. For this study, the N-gram size was set to four to identify four-word LBs, and the minimum range for each LB to occur was set at three.

To ensure the reliability of the data coding for the structural and functional classification of the identified LBs extracted from AntConc, the first and second authors together coded all extracted LBs. Rather than calculating percentage agreement, the two coders discussed any discrepancies and negotiated until consensus was reached for all identified LBs, meaning that all the classifications were agreed by both coders.

It should be noted that LBs do not always constitute complete structural units with semantic or figurative meanings (e.g., Biber & Barbieri, 2007; Panthong & Poonpon, 2020). However, all of the identified LBs were taken into consideration for analysis in this study.

RESULTS AND DISCUSSION

The 30 oral presentations delivered by the participants in the present corpus represented a form of specialized academic discourse, characterized by the structural and functional use of LBs. As Hyland (2009) noted, undergraduate students' oral presentations are a valuable area of inquiry since they typically exemplify an "informal, audience-friendly presentation" (p. 139). This section presents the list of four-word LBs identified in the corpus, categorized according to their functional distribution and structural patterns following Biber et al. (2004).

Functional distribution of lexical bundles

Table 3 below presents the functional distribution of four-word LBs identified in the current corpus. The analysis identified 284 instances of LBs and 58 distinct ones in the corpus. They were categorized into four major functional types (i.e., stance expression, discourse organizers, referential expression, and special conversational functions) following Biber et al.'s (2004) taxonomy. Each function was further divided into sub-categories, and where applicable, sub-subcategories.

Table 3
Functional distribution of four-word lexical bundles (frequency & percentage) (N = 284)

Functional Type	Subcategory	Sub-subcategory	(N = 284) Frequency	Percentage N = 284	Distinct LBs	Example
Stance expressions	Epistemic stance		44	15.49%	9	my opinion I think
	Attitudinal/ modality stance	Desire	5	1.76%	1	I would like to
		Obligation/ directive	-	-	-	-
		Intention/ prediction	3	1.06%	1	that I’m going
		Ability	-	-	-	-
	Total of the subcategory		52	18.31%	11	
Discourse organizers	Topic introduction/focus		133	46.83%	22	going to present about
	Topic elaboration/clarification		20	7.04%	5	as you can see
	Total of the subcategory		153	53.87%	27	-
Referential expressions	Identification/focus		9	3.17%	2	is one of the
	Imprecision		15	5.28%	4	pros and cons and
	Specification of attributions	Quantity specification	-	-	-	-
		Tangible attributes	-	-	-	-
		Intangible attribution	-	-	-	-
	Time/ place/text reference	Place ref.	-	-	-	-
		Time ref.	-	-	-	-
		Text deixis	-	-	-	-
		Multifunctional reference	3	1.06%	1	at the end of
	Total of the subcategory		27	9.51%	7	
Special conversational functions	Politeness		35	12.32%	9	thank you for listening
	Simple inquiring		17	5.99%	4	if you have any
	Reporting		-	-	-	-
	Total of the subcategory		52	18.31%	13	-
	Total		284	100%	58	

Stance expression

As shown in Table 3 above, the analysis revealed 58 distinct LBs with a total of 284 occurrences identified according to the criteria of the present study, following Biber et al.'s (2004) taxonomy, which classifies bundles into three main functions. Stance expression bundles were the first category used to present speakers' attitudes or evaluations in relation to the degree of certainty where they frame a given proposition (Biber et al., 2004). A total of 52 occurrences (18.31%) of stance expression bundles were identified across the corpus, distributed across the two subcategories (i.e., epistemic and attitudinal/modality).

As seen in Table 3 above, epistemic stance bundles appeared 44 times (15.49%). They are used to indicate the speaker's epistemic stance, presenting personal opinions and hedging the claims. For instance, in excerpt (1) below, the presenter mitigated the claim using the bundle *my opinion, I think*, illustrating how the speaker toned down the degree of certainty and presented the viewpoint cautiously.

In *my opinion, I think* it will be fundamental for the future. (1)

Comparing the current findings with related studies, Shin and Won (2024) found the prevalence of epistemic stance bundles in a corpus of English argumentative speeches delivered by Korean EFL learners. This reflected the argumentative nature of the discourse. However, the current findings slightly differ from those of Shin and Won (2024), as stance bundles were found to be the second most common type. This may be due to the nature of the topics the students presented, as they aimed to provide readers with information rather than to advance specific claims.

Following epistemic bundles, another subcategory under stance expression identified was an attitudinal/modality stance. Only desire and intention/prediction bundles were found in the corpus, and their rare occurrence leads to some interesting points for discussion. On the one hand, it can be argued that this type of bundle is not a commonly used one in students' oral presentations. On the other hand, it may also reflect that students lack the ability to articulate such functional bundles to express their attitudes or intentions effectively.

Excerpts 2 and 3 below illustrate the use of desire and intention/prediction bundles in the current corpus. The undergraduate students tended to rely only on the bundle *I would like to* when they aimed to introduce or initiate a topic, reflecting a desire to share information with the audience while the bundle *I'm going to* was used as a building block to signal to the audience the upcoming content. After these two bundles are employed, the audience can anticipate either the introduction of a new topic or an explanation of the points that will follow.

I would like to talk about ergonomic chairs. (2)

...this technology that *I'm going to* tell you which is called cryogenic sleep... (3)

The use of desire and intention/prediction bundles might be due to their functions, which are to express personal wishes, preferences, or anticipated actions rather than to present

information academically. Biber and Barbieri (2007) found a greater number of desire bundles in a corpus of spoken classroom management as they can be directly used to inform the students what the teacher ‘wants’ or ‘would like’ (p. 281).

Discourse organizer

It was found that the most dominant type of LBs is discourse organizers with 153 occurrences (53.87%) across 27 distinct bundles. Within this category, the topic introduction/focus subcategory was the most frequent, with 133 occurrences (46.83%), followed by topic elaboration/clarification, with 20 occurrences (7.04%).

Evidence of the use of the topic introduction/focus function can be observed as a form of syntactic blending (Biber et al., 1999; Biber et al., 2004). This linguistic pattern occurs when two parts of the discourse are combined (i.e., an LB followed by the new topic it introduces). As illustrated in excerpts 4 and 5 below, the bundle *going to present about* was used to precede the topic to be presented, that is, *processing technologies to decrease the risk*, effectively signaling the introduction of a new topic in former, while the latter with the bundle *going to talk about* precedes *technology in mental health*. In both excerpts, these topic-introducing bundles were used to inform the audience with the upcoming topic in students’ oral presentations.

Today I’m *going to present about* processing technologies to decrease the risk. (4)

Today I’m *going to talk about* the topic technology in mental health. (5)

Five different LBs contained the word move: *let’s move on*, *’s move on to*, *move on to the*, *let’s move to*, and *so let’s move*. This echoes Kwon and Lee (2014), who found that bundles containing the phrasal verb *move on* were more frequently used by non-native speakers of English compared to native speakers. To illustrate this, in the current corpus the bundle *let’s move on* was employed in excerpt (6) below to signal a transition to the next part of the presentation, *the pros and cons*. This suggests that non-native speakers of English appear to have a limited range of expressions for topic elaboration/clarification, as they relied heavily on the phrasal verb *move on* in the current corpus.

And *let’s move on* to the next part, the pros and cons of this cryogenic sleep. (6)

The second subcategory of LBs under discourse organizers was topic elaboration/clarification, which occurred 20 times across five distinct types, such as, *as you can see*, and *you can see that*. This suggests that the students could employ bundles to elaborate on and clarify their ideas. Such bundles function to provide additional explanations and guide the audience’s understanding of the presentation. The use of these bundles can also signal to the audience that the speaker is providing benefits, explanations or noting significance. As illustrated in excerpt (7) below, a topic elaboration/clarification bundle was used to provide detailed explanation, in which the listeners were provided with additional information about the significance of an ergonomic chair.

You can see that an ergonomic chair will improve your culture, support your back. (7)

Referential expressions

The third category was referential, which enables speakers to point out either physical or abstract references. It can also function to refer backward or forward within the text itself (Biber & Barbieri, 2007). Identification/focus bundles were found nine times with two different bundles while imprecision bundles were found 15 times with four different bundles. The analysis revealed only one multifunctional reference bundle under the time/place/text reference. With its rare occurrence, it can be argued that referential expression may not be commonly used in undergraduate presentations. A sample of the bundle *at the end of* is exemplified in excerpt (8) below.

...but it was a football game for me sometime when was analyzed at almost *at the end of the game*. (8)

Previous studies revealed the occurrence of referential expressions in spoken discourse. Zipagan and Lee (2018) found that advanced EFL Korean learners employed more referential expressions than their intermediate and novice counterparts. Heng et al. (2014) found the use of referential expressions in group discussions. The scant identification of referential expressions in the current corpus may be attributed to the differences in genres and task demands, given that oral presentations tend to deliver information in a limited time given. Therefore, classroom teachers should encourage students to incorporate referential expressions by designing oral presentation tasks that require explicit reference to data.

Special conversational functions

It was evident that students employed a wide range of special conversational bundles. In particular, politeness bundles appeared 35 times with nine distinct bundles, while simple inquiring bundles were found 17 times with four different ones. Excerpt (9) below demonstrates the use of the politeness strategy through the bundle *thank you for listening*, whereas the use of a simple inquiring bundle signals that the speaker is open for interaction with the audience though the use of *If you have any* in excerpt (10) below.

Yeah, *thank you for listening*. (9)

If you have any questions about my presentation, you can ask me back for this. (10)

A possible explanation may be attributed to pedagogical practices in the classroom of the current study. The students were taught the conventional structure of oral presentations (i.e., introduction, body, and conclusion). They were also advised to begin with greetings and conclude by expressing gratitude to the audience. Accordingly, the bundles *if you have any* signals to the audience that the speaker is open to questions, comments, or interaction, which could enhance engagement and dialogue at the end of the presentation.

The findings regarding the functional distribution of LBs in the current corpus have led to several points for discussion. Firstly, the analysis indicated that discourse organizers emerged as the most frequently identified functional category of LBs in the present corpus, accounting

for 53.87%. The findings reflect the nature of oral presentations in which speakers need to signal or introduce new topics to the audience. In contrast, a study on LBs used in a speaking test by Korean EFL learners found stance expressions as the most dominant type (Zipagan & Lee, 2018). Secondly, the absence of certain functional subcategories may be due to the nature of the oral presentation task itself in the current context. For example, since the task did not require the presenters to provide detailed explanations, such as numerical data related to the topic being presented, the subcategory of specification of attributes did not emerge in the corpus. However, this subcategory was frequently employed in group discussions in the context of university talk (Heng et al., 2014).

This difference may explain why the present findings contrast with those of Heng et al. (2014), who reported the common use of specification of attributes to express quantity (e.g., *a little bit*) in group discussions. Classroom teachers may explicitly teach students to incorporate numerical data in their presentation not only to make the presentation more engaging but also to allow them to articulate specifications of attribution LBs. Another noteworthy point is the frequent use of certain LBs, which could indicate a limited spoken repertoire among the presenters. In particular, bundles containing the word *going*, such as *I am going to*, *today I am going*, and *going to talk about*, occurred repeatedly in the corpus. These findings contribute to important pedagogical implications, suggesting that teachers should incorporate a wider variety of LBs into their oral presentation instruction and teaching materials in order to expand students' spoken repertoire and reduce overreliance on a limited set of LBs.

Structural distribution of lexical bundles

The second criterion for the extraction of bundles in the current study was structural patterns, following Biber et al. (2004). Table 4 below presents the structural distribution of four-word LBs identified in the current corpus. Overall, three main structural types were identified, including (1) verb phrase fragments, (2) dependent clause fragments, and (3) noun phrase and prepositional phrase fragments.

Table 4
Structural distribution of four-word lexical bundles

Main Structural Type	Subcategory	Frequency	Percentage (N = 284)	Numbers of different LBs	Example Bundles
1. Lexical bundles that incorporate verb phrase fragments	1a. (Connector) + 1st/2nd pronoun + VP fragment	99	34.86%	16	I'm going to today I'm going
	1b. (Connector) + 3rd pronoun + VP fragment	54	19.01%	11	the last one is
	1c. Discourse marker + VP fragment	-	-	-	-
	1d. Verb phrase with non-passive verb	50	17.61%	11	going to talk about
	1e. Verb phrase with passive verb	-	-	-	-

Main Structural Type	Subcategory	Frequency	Percentage (N = 284)	Numbers of different LBs	Example Bundles
	1f. Yes–no question fragments	5	1.76%	1	you have any question
	1g. WH-question fragments	4	1.40%	1	how does it work
	Total of the subcategory	212	74.65%	40	
2. Lexical bundles that incorporate dependent clause fragments	2a. 1st/2nd pronoun + dependent clause fragment	9	3.17%	1	my opinion I think
	2b. WH-clause fragments	0		0	
	2c. If-clause fragments	8	2.82%	2	if you have any
	2d. Verb/adjective + to-clause fragment	5	1.76%	1	I would like to
	2e. That-clause fragments	3	1.06%	1	that I'm going
	Total of the subcategory	25	8.80%	5	
3. Lexical bundles that incorporate noun phrase and prepositional phrase fragments	3a. (Connector) + NP with of-phrase	7	2.46%	2	advantage and disadvantage of pros and cons of
	3b. NP with other post-modifier	3	1.06%	1	the end of the
	3c. Other NP expressions	17	5.99%	4	last part is my and the next one
	3d. Prepositional phrase expressions	16	5.63%	4	in my opinion I
	3e. Comparative expressions	4	1.41%	1	one of the most
	Total of the subcategory	47	16.55%	12	
Grand total		284	100		

Lexical bundles that incorporate verb phrase fragments

There were 212 LBs with verb phrase fragments, making this the most dominant structural type in the corpus. The largest subcategory was bundles with connector + 1st/2nd pronoun + VP fragment, such as *I'm going to* or *today I'm going*, which occurred 99 times across 16 different bundles. Excerpt (11) below demonstrates the use of *I'm going to* in the corpus.

Today *I'm going to* present about processing technologies to decrease the risk. (11)

As the second most frequent subcategory with verb phrase fragments, the Connector + 3rd pronoun + VP fragment bundle, such as *the last one is* was found 54 times across 11 bundles, as illustrated in excerpt (12) below.

... and *the last one is* my opinion. (12)

Verb phrase with non-passive verb bundles occurred 50 times across 11 different bundles, as shown in excerpt (13) below with the use of *going to talk about*.

Today I'm *going to talk about* the Human Genome Project. (13)

The bundle, *you have any question*, was the only bundle with yes–no question fragments, appearing five times as shown in excerpt (14).

And if *you have any question* of my presentation... (14)

The bundle, *how does it work*, was the only one with WH-question fragments, appearing four times, as in excerpt (15) below.

Let's turn to the next part. *How does it work?* (15)

In the Thai Context, LBs incorporating verb phrase fragments were found to dominate both spoken and written academic discourse. In the latter, Leelasetakul (2019) found that Thai EFL freshmen employed LBs incorporating verb phrase fragments most frequently in their argumentative essays, highlighting their reliance on this structural pattern. A similar reliance was similarly identified in the current study, suggesting that verb phrase fragments also play a crucial role in spoken academic discourse. Such bundles help students structure their talk and signal transitions in oral presentations. The identification of bundles incorporating verb phrase fragments also echoes Conrad and Biber (2005), who found that 90% of bundles in academic discourse are composed of verb phrase fragments. The findings echo L2 users of LBs by Bychkovska and Lee (2017) who found that L2 learners of English tended to employ significantly more bundles with verb phrase more than L1 learners. The findings of the heavy use of LBs with verb phrase fragments in the current corpus point out that not only do the findings reflect the role of VP fragments in managing oral presentations, but they also reflect the characteristics of L2 learners of English.

Lexical bundles that incorporate dependent clause fragments

The second structural category was LBs that incorporated dependent clause fragments. Twenty-five occurrences across five different bundles were found, making this category less frequent than verb phrase fragments in students' oral academic discourse. The most frequent subcategory was 1st/2nd pronoun + dependent clause fragment, accounting for nine times through the use of *my opinion I think*, indicating students' attempt to pinpoint personal stance from a subjective perspective as seen in excerpt (16) below.

In *my opinion, I think* it will be the fundamental for the future. (16)

The second most common type was if-clause fragments, appearing eight times across two bundles: *and if you have* and *if you have any*. They were often used to invite audience questions and maintain interaction at the end of presentations, as seen in excerpt (17) below.

And if you have a question, you can ask me. (17)

Bundles with verb/adjective + to-clause fragments occurred five times with *I would like to* as the single bundle in the corpus used to politely indicate intentions, contributing to a more formal and respectful presentation style, as illustrated in excerpt (18) below.

... what *I would like to* say to you today... (18)

That-clause fragment bundles were identified three times though the use of *that I'm going*. This structural bundle occurred when the speakers aimed to extend the preceding statements. Although relatively infrequent, this structure reflects an effort to integrate more complex syntactic patterns into oral discourse, as shown in excerpt (19) below.

...and these are contents *that I'm going* to cover today. (19)

The identification of 25 bundles with dependent clause fragments indicated that the students tended to translate directly from Thai into English. To illustrate this, the bundle *that I'm going* functions as part of a relative clause modifying a preceding noun (e.g., *the topics*) as seen in excerpt (19) above. Although not as frequent as verb phrase fragments, these bundles revealed that learners could enhance cohesion and interaction in academic presentations. Dependent clauses also played a vital role in elaborating on the points that students would like the audience to understand. The findings are similar to the discourse of lecturers in hard sciences as suggested by Kashiha and Heng (2013).

Lexical bundles that incorporate noun phrase and prepositional phrase fragments

The last structural category identified in the corpus was bundles with noun phrase and prepositional phrase fragments. A total of 47 tokens with 12 different bundles were found. It was evident that these bundles reveal students' strategies for marking transitions, signaling structure, and organizing ideas within oral presentations.

Bundles with connector + NP with of-phrase occurred seven times across two bundles. The bundles *pros and cons of* and *advantage and disadvantage of* were used by students to introduce points of comparison or evaluation, often signaling both sides of the presentation. These two bundles constituted key lexical resources in the students' oral presentations, in which they presented balanced arguments or evaluations of their selected topic, as shown in excerpt (20) below.

...the *advantage and disadvantage of* the Invisalign... (20)

The function of these bundles was mainly to quantify or identify the two sides of the topic being discussed. Kashiha and Heng (2013) also found that this kind of structural bundle was frequently used in classroom lectures to quantify, qualify, or identify key points. The current findings also echo Kashiha and Heng (2013), who reported that bundles containing noun phrase with of-phrase fragments were the dominant type of bundles with noun phrase and prepositional phrase fragments.

Bundles with NP + other post-modifier appeared three times through the use of *the end of the*. This bundle was employed to mark the specific point referent to the time or stage of a topic being discussed as shown in excerpt (21) below.

.... analyzed at almost at *the end of the* game. (21)

Bundles with other NP expressions occurred 17 times across four bundles, such as *last part is my*, making them the most common structural pattern within this sub-category. These bundles generally served to signal a shift or transition in the presentation, helping the audience follow the organization of the talk. For example, excerpt (22) below demonstrates this type of bundle being used to indicate that the presenter was about to express his or her opinion at the end of the presentation.

.... and the *last part is my* opinion. (22)

Bundles with prepositional phrase expressions appeared 16 times across four bundles. A common example was *in my opinion I*, which reflected students' attempts to pinpoint their stance and clarify their personal evaluation or tone down the claims as seen in excerpt (23).

...*in my opinion*, I think using... (23)

One bundle with a comparative expression was found in the corpus. In excerpt (24), *one of the most* was the only example used to indicate emphasis or ranking among features or points, suggesting that students rarely employed comparative sentences in their oral presentations.

ADHD is *one of the most* common developmental diseases of childhood. (24).

A study by Bychkovska and Lee (2017) indicated that L1 learners employed more noun and prepositional phrase bundles than L2 Chinese learners in English argumentative essays. The current findings appear to align with Bychkovska and Lee's (2017) results, as noun and prepositional phrase bundles in the present study were relatively less frequent than other structural patterns. This reflects the nature of spoken discourse among L2 learners of English, in which non-native speakers tend to rely more on verb phrase-based expressions to organize their speech and convey ideas fluently.

To sum up, the findings provided insights into the patterns of LBs in Thai EFL oral presentations, in which verb phrase fragments were the most prevalent structural type, accounting for 74.65% of the corpus. This predominance reflects the nature of oral presentations, which tend to rely heavily on verb-based constructions to maintain the flow of spoken discourse. These findings are consistent with previous studies as verb phrase fragments were the dominant structural type of LBs in spoken discourse, such as speaking test performances (Zipagan & Lee, 2018) and conversation (Biber & Conrad, 2005). This may be due to the nature of spoken discourse, in which verbs play a crucial role in expressing actions with the use of the personal pronoun, such as *I don't know what* (Biber & Conrad, 2005).

However, the heavy reliance on verb-based fragments, such as *I'm going to* and *today I'm going to*, may also be the characteristic of EFL learners. For instance, Bychkovska and Lee (2017) compared the use of LBs by native speakers and Chinese EFL learners in argumentative essays and found that the latter employed more verb phrase LBs than the former. The identification of this heavy reliance on verb phrase fragments in the current corpus suggests that students should be exposed to a wider range of structural types of LBs in order to be capable of articulating a wider range of LBs in oral presentations.

CONCLUSION

The present study analyzed four-word lexical bundles in 30 undergraduate oral presentations, revealing important insights into the functional and structural patterns of spoken academic discourse among Thai EFL learners. Overall, following Biber et al.'s (2004) taxonomy, a total of 284 occurrences across 58 distinct bundles were identified and categorized according to their functions and structures. Discourse organizers were found to be the most dominant functional category, indicating the students' strong reliance on formulaic sequences to manage information flow and structure their speech. Bundles with verb phrase fragments were the most frequently employed structural pattern. On the one hand, the findings indicate the reliance and tendency of students to use LBs to deliver and organize their oral presentations effectively. On the other hand, the absence of several functional and structural types suggests a limited repertoire of academic formulaic language, which may call for more explicit instruction and practice in using a wider range of LBs to enhance fluency and academic expression in spoken discourse.

It should be noted that the current study is not without limitations. The first limitation is the relatively small corpus size of 30 oral presentations. This could restrict the generalizability of the findings. Secondly, the data were drawn from a single academic context. The identified LBs may not fully represent the broader use of LBs by Thai EFL undergraduates. Future studies may consider expanding the corpus size and collecting data from a wider range of academic settings to provide a more comprehensive understanding of LB use.

ACKNOWLEDGEMENT

This study was financially supported by Navamindradhiraj University Research Fund, Navamindradhiraj University (Contract No. 004/2568).

THE AUTHORS

Soranut Kumdee is an English instructor at the Department of General Education, Faculty of Sciences and Health Technology, Navamindradhiraj University, Thailand. He earned his Ph.D. in English Language Studies at the School of Foreign Languages, Suranaree University of Technology. His research interests include genre analysis, professional discourse, sports news discourse, specialized genres, learner corpus and corpus linguistics.

soranut.k@nmu.ac.th

Kanokpoj Khakhai is an English instructor at the Department of General Education, Faculty of Sciences and Health Technology, Navamindradhiraj University, Thailand. His research interests cover English for specific purposes, needs analysis, language teacher education, teacher beliefs, practices, and identity.

kanokpoj.k@nmu.ac.th

REFERENCES

- Anthony, L. (2024). *AntConc* (Version 4.3.1) [Computer Software]. Waseda University. <https://www.laurenceanthony.net/software/AntConc>.
- Bankowski, E. (2010). Developing skills for effective academic presentations in EAP. *International Journal of Teaching and Learning in Higher Education*, 22(2), 187–196.
- Barbieri, F. (2018). I don't want to and don't get me wrong: Lexical bundles as a window to subjectivity and intersubjectivity in American blogs. In J. Kopaczuk & J. Tyrkkö (Eds.), *Applications of pattern-driven methods in corpus linguistics* (pp. 251–275). John Benjamins. <https://doi.org/10.1075/scl.82.10bar>
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263–286. <https://doi.org/10.1016/j.esp.2006.08.003>
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied linguistics*, 25(3), 371–405. <https://doi.org/10.1093/applin/25.3.371>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.
- Bowden, M. (2023). *How to present: The ultimate guide to presenting live and online* (2nd ed.). Wiley.
- Brooks, G., & Wilson, J. (2014). Using oral presentations to improve students' English language skills. *Kwansei Gakuin University Humanities Review*, 19(1), 199–212.
- Bychkovska, T., & Lee, J. J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes*, 30, 38–52. <https://doi.org/10.1016/j.jeap.2017.10.008>
- Chinapakdee, M. (2015). Thai EFL university students' perspectives on foreign language anxiety. *Humanities, Arts and Social Sciences Studies*, 42, 61–90.
- Conrad, S. M., & Biber, D. (2005). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica*, 20, 56–70. <https://doi.org/10.1515/9783484604674.56>
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397–423. <https://doi.org/10.1016/j.esp.2003.12.001>
- Cortes, V. (2015). Situating lexical bundles in the formulaic language spectrum. In V. Cortes & E. Csomay (Eds.), *Corpus-based research in applied linguistics* (pp. 197–216). John Benjamins.
- Dan, C., Darmi, R. H., & Yusof, M. A. M. D. (2024). The use of lexical bundles in English language academic writing among university learners: A systematic literature review. *World Journal of English Language*, 14(3), 196–210. <https://doi.org/10.5430/wjel.v14n3p196>

- Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 42(3), 375–396. <https://doi.org/10.1002/j.1545-7249.2008.tb00137.x>
- Farhang-Ju, M., Jalilifar, A., & Keshavarz, M. H. (2024). Specificity and generality of lexical bundles in the rhetorical moves of applied linguistics research article introductions. *Journal of English for Academic Purposes*, 69, Article 101387. <https://doi.org/10.1016/j.jeap.2024.101387>
- Farrington, J. (2022). *The non-obvious guide to better presentations: How to present like a pro (virtually or in person)*. Ideapress.
- Flinn, A. (2023). How often do pauses occur in lexical bundles in spoken native English speech? *Corpus Pragmatics*, 7, 303–322.
- Gong, H., Le, T. N. P., & Buckingham, L. (2025). Lexical bundles across IMRD-structured Medicine research article sections: A within-register perspective. *Journal of English for Academic Purposes*, 74, Article 101487. <https://doi.org/10.1016/j.jeap.2025.101487>
- Heng, C. S., Kashiha, H., & Tan, H. (2014). Lexical bundles: Facilitating university “talk” in group discussions. *English Language Teaching*, 7(4), 1–10. <https://doi.org/10.5539/elt.v7n4p1>
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4–21. <https://doi.org/10.1016/j.esp.2007.06.001>
- Hyland, K. (2009). *Academic discourse: English in a global context*. Continuum.
- Kalra, R., & Siribud, S. (2020). Public speaking anxiety in the Thai EFL context. *LEARN Journal: Language Education and Acquisition Research Network*, 13(1), 195–209.
- Kashiha, H., & Heng, C. S. (2013). An exploration of lexical bundles in academic lectures: Examples from hard and soft sciences. *The Journal of Asia TEFL*, 10(4), 133–161.
- Kurakan, P. (2021). Anxiety in English oral presentations of Thai EFL engineering students. *THAITESOL Journal*, 34(2), 67–92.
- Kwon, Y. E., & Lee, E. J. (2014). Lexical bundles in the Korean EFL teacher talk corpus: A comparison between non-native and native English teachers. *Journal of Asia TEFL*, 11(3), 73–103.
- Laosrirattanachai, P., & Laosrirattanachai, P. (2024). Hope you guys enjoyed this Vlog: Lexical bundles and moves within English travel Vlog discourse. *LEARN Journal: Language Education and Acquisition Research Network*, 17(2), 249–278. <https://doi.org/10.70730/VIZO4161>
- Leelasetakul, M. (2019). *Lexical bundles use by Thai undergraduate learners of English* [Unpublished doctoral dissertation]. Pennsylvania State University.
- Li, L., Franken, M., & Wu, S. (2020). Bundle-driven move analysis: Sentence initial lexical bundles in PhD abstracts. *English for Specific Purposes*, 60, 85–97. <https://doi.org/10.1016/j.esp.2020.04.006>
- Limgomolvilas, S. (2024). Differentiating learner’s English proficiency in oral presentation focusing on textual markers. *rEFlections*, 31(2), 501–515. <https://doi.org/10.61508/refl.v31i2.274803>
- Lowe, S., & Pile, L. (2006). *Presenting*. Delta.
- Narkprom, N., & Phoocharoensil, S. (2022). Lexical bundles in native English speakers’ and Thai writers’ dissertations. *GEMA Online Journal of Language Studies*, 22(3), 137–152.
- Naser, N. A. M., & Isa, I. A. M. (2021). Public speaking anxiety in oral presentation class among undergraduates. *International Journal of Academic Research in Business and Social Sciences*, 11(10), 877–889.
- Öztürk, Y., & Taşçı, S. (2023). A corpus-based analysis of lexical bundles in non-native post graduate academic writing and a potential L1 influence. *rEFlections*, 30(2), 488–505. <https://doi.org/10.61508/refl.v30i2.267463>
- Panthong, P., & Poonpon, K. (2020). Functional analysis of lexical bundles in doctor talks in the medical TV series Grey’s Anatomy. *LEARN Journal: Language Education and Acquisition Research Network*, 13(2), 335–353.

- Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 1–12). John Benjamins.
- Shin, Y. K., & Won, D.-O. (2024). To what extent do L2 learners produce genre-appropriate language? A comparative analysis of lexical bundles in argumentative essays and speeches. *Journal of English for Academic Purposes*, 69, Article 101389. <https://doi.org/10.1016/j.jeap.2024.101389>
- Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. Routledge.
- Siricharoen, A., & Wijitsopon, R. (2020). A corpus-based comparative study of lexical bundles in authentic and textbook English business emails. *LEARN Journal: Language Education and Acquisition Research Network*, 13(2), 41–63.
- Smakman, D. (2024). *Presenting in English: A practical guide*. Routledge. <https://doi.org/10.4324/9781032627892>
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. John Benjamins.
- Wang, W., & Csomay, E. (2024). Constructing proximity in popularization discourse: Evidence from lexical bundles in TED talks. *English for Specific Purposes*, 73, 95–109. <https://doi.org/10.1016/j.esp.2023.10.003>
- Wray, A. (2013). Formulaic language. *Language Teaching*, 46(3), 316–334. <https://doi.org/10.1017/S0261444813000013>
- Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language & Communication*, 20(1), 1–28. [https://doi.org/10.1016/S0271-5309\(99\)00015-4](https://doi.org/10.1016/S0271-5309(99)00015-4)
- Wright, H. R. (2019). Lexical bundles in stand-alone literature reviews: Sections, frequencies, and functions. *English for Specific Purposes*, 54, 1–14. <https://doi.org/10.1016/j.esp.2018.09.001>
- Zareva, A. (2024). Effectiveness of student academic presentations: What do we learn from presentation guidebooks? *Journal of English for Academic Purposes*, 71, Article 101423. <https://doi.org/10.1016/j.jeap.2024.101423>
- Zipagan, M. N., & Lee, K. R. (2018). Korean English learners' use of lexical bundles in speaking. *Journal of Asia TEFL*, 15(2), 276–291.

Appendix A

A list of lexical bundles identified

Bundle	Rank	Freq	Range
I'm going to	1	22	17
today I'm going	2	12	12
as you can see	3	9	4
my opinion I think	3	9	9
going to talk about	5	8	7
m going to talk	5	8	7
the last one is	7	7	6
and the last one	8	6	6
m going to present	8	6	6
part is my opinion	8	6	5
the next one is	8	6	4
the next part is	8	6	3
everyone today I m	13	5	5
going to present about	13	5	5
I would like to	13	5	4
if you have any	13	5	4
in my opinion I	13	5	5
is one of the	13	5	4
last part is my	13	5	4
let s move on	13	5	4
pros and cons and	13	5	5
s move on to	13	5	4
the first one is	13	5	4
the last part is	13	5	4
you have any question	13	5	3
advantage and disadvantage of	26	4	3
afternoon everyone today I	26	4	4
and the next one	26	4	3
good afternoon everyone today	26	4	4
hello everyone today I	26	4	4
hello my name is	26	4	4
how does it work	26	4	3
is my opinion I	26	4	4
let's start with	26	4	3
move on to the	26	4	4
my opinion in my	26	4	4
one of the most	26	4	4
opinion I think this	26	4	4
opinion in my opinion	26	4	4

Bundle	Rank	Freq	Range
so this is the	26	4	4
thank you for listening	26	4	4
thank you for your	26	4	4
to the next part	26	4	3
and if you have	46	3	3
and the last is	46	3	3
and the last part	46	3	3
everyone my name is	46	3	3
I think it s	46	3	3
let's move to	46	3	3
on to the next	46	3	3
pros and cons of	46	3	3
so let s move	46	3	3
that I'm going	46	3	3
the advantage and disadvantage	46	3	3
the end of the	46	3	3
you can see that	46	3	3
you for your attention	46	3	3
Total		284	