Formulaic Language Use by Learners of English in Interlanguage Communication

Zhaoyi Pan[1]

[1] English Language Coordinating Section, School of Information Technology, King Mongkut's University of Technology Thonburi, Bangkok, Thailand

Correspondence: Zhaoyi Pan, English Language Coordinating Section, School of Information Technology, King Mongkut's University of Technology Thonburi, 126 Pracha-Uthit Rd., Bangmod, Thungkhru, Bangkok 10140, Thailand.

Received: May 9, 2023 Accepted: June 12, 2023 Online Published: July 11, 2023

doi:10.5430/wjel.v13n7p128 URL: https://doi.org/10.5430/wjel.v13n7p128

Abstract

This study investigated the use of formulaic language, in the form of four-word lexical bundles, by Thai learners of English as a second language (ESL) at various levels of English proficiency during interlanguage communication. The investigation focused on two aspects: the frequencies and pragmatic functions of the four-word lexical bundles. A total of 120 Thai ESL learners participated in the study, ranging from basic to intermediate and advanced English proficiency levels. In terms of frequency, a list of the most frequently used four-word lexical bundles by Thai ESL learners at each level was examined. Results showed that learners at higher English levels used less formulaic language in interactions than those at lower levels. Two similar four-word lexical bundles, centered on "I don't know" and "I think," were used by Thai ESL learners at all three English levels. The functional analysis demonstrated that Thai ESL learners at all three English levels used formulaic language to assert group identity as a device for social interaction and as a device for memory limitations to buy more time and process shortcuts. The function of asserting a separate identity was not used by intermediate-level Thai ESL learners. Furthermore, the function of processing shortcuts was used less at the advanced level than at the basic level. The findings of this study indicate that Thai ESL learners use formulaic language differently at different English proficiency levels. Therefore, formulaic language used by ESL learners in spoken English discourse should be further investigated.

Keywords: interlanguage pragmatics, formulaic language, four-word lexical bundles, Thai ESL learners, pragmatic functions

1. Introduction

Previous studies on formulaic language have demonstrated that native English speakers primarily use fixed or non-fixed lexical units stored in memory in both written and spoken English (Wray, 2002, 2008; Wray & Perkins, 2000). According to Erman and Warren (2000), formulaic language makes up 58.6% of spoken English and 52.3% of written English, highlighting the significance and prevalence of formulaic language, which challenges the traditional notion that language consists solely of grammar and a single lexicon (Wray & Perkins, 2000).

Recently, researchers have become interested in the use of formulaic language by learners of English as a second language (ESL; Ebrahimi et al., 2021; Wang & Halenko, 2022). Studies have extensively examined the use of formulaic language in academic English writing (Chen & Baker, 2016; Leelasetakul, 2019; Narkprom & Phoocharoensil, 2022; Staples et al., 2013) and the effects of different teaching methods on developing formulaic language use by ESL learners (Ogawa, 2021; Rajeswaran, 2019). However, Staples and Fernández (2019) found that formulaic language has not been adequately investigated in relation to interlanguage pragmatics, particularly in spoken English discourse. Thus, this study examines the frequency and pragmatic functions of formulaic language used by Thai ESL learners in spoken English discourse. The findings reveal the patterns of formulaic language used by ESL learners in spoken English discourse.

2. Literature Review

2.1 Formulaic Language

Formulaic language has been described in various terms in the literature (Wray, 2002), such as fixed or semi-fixed units (Biber, 2009), formulaic sequences (Alharthi, 2015), lexical bundles (Leelasetakul, 2019), formulaic expressions (Ebrahimi et al., 2021), phrase frames (Tan & Römer, 2022), and multiword structures (Erman & Lewis, 2022). However, previous research largely agrees with the concept of formulaic language proposed by Wray (2002), who defines formulaic language as "a sequence of words, either continuous or discontinuous" (p. 9). Additionally, formulaic language appears to be prefabricated in the memory of language users. As a result, it can be retrieved as an entity while being produced (Wray & Perkins, 2000). This definition reveals two aspects of formulaic language. First, formulaic language consists of several single lexical items, either continuous (fixed) such as "on the other hand" (Leelasetakul, 2019), or discontinuous (semi-fixed) such as "the * of the" (Biber, 2009). Second, these combinations of single lexical items are not random collocations, but instead, they are memorized by language users as prefabricated combinations and used by them when required in communication.

Studies on formulaic language have focused on both native English speakers and ESL learners. The most frequently used formulaic language has always been of interest in this field. Several studies have identified the linguistic forms of the most frequently used formulaic language (Bardovi-Harlig & Stringer, 2017; Chen & Baker, 2014; Erman & Lewis, 2022), such as noun phrases and verb phrases. Moreover, the various pragmatic functions of formulaic language in both written and spoken discourse have been evaluated (Narkprom & Phoocharoensil, 2022; Staples et al., 2013; Staples & Fernández, 2019). Specifically, the pragmatic functions of formulaic language in academic English writing have been analyzed (Leelasetakul, 2019), mostly based on the frameworks proposed by Biber (2009) and Hyland (2008). It was found that native English speakers use formulaic language as stance expressions more frequently than ESL learners (Leelasetakul, 2019), while ESL learners use formulaic language as text organizers at a high frequency (Narkprom & Phoocharoensil, 2022).

2.2 Pragmatic Functions of Formulaic Language

Formulaic language plays a significant role in pragmatics because of its pragmatic value in discourses (Wray & Perkins, 2000). Language users in the same community often use the same or similar recurrent communicative expressions to align with each other and minimize the risk of potential misunderstandings (Wray, 2008). Previous research (Narkprom & Phoocharoensil, 2022) found that, in academic English writing, authors express their stances to connect with readers and avoid misunderstandings by using different formulaic language.

The formulaic language utilized in spoken discourse, similar to that in academic English writing, possesses its own pragmatic value (Wray, 2002). To investigate the pragmatic functions of formulaic language in speech, the frameworks proposed by Biber (2009) or Hyland (2008) may not be optimal, as they were developed and are primarily employed for written language (Narkprom & Phoocharoensil, 2022; Staples et al., 2013). In contrast, Wray and Perkins' (2000) framework for the pragmatic functions of formulaic language was specifically designed for the study of formulaic language in speech and was employed in this study.

In Wray and Perkins' (2000) framework, the pragmatic functions of formulaic language in speech are classified into two fundamental types: devices for social interactions and devices to compensate for memory limitations in interaction, as presented in Table 1.

Table 1. Pragmatic functions of formulaic language in spoken discourses adapted from Wray and Perkins (2000)

Pragmatic functions of formulaic language			
Devices for social interactions		Devices for memory limitations	
Manipulation of others	Commands	Processing short-cuts	Standard phrases (with or without gaps)
	Requests		
	Politeness markers		Standard ideational labels with agreed meaning
	Bargains		
Asserting separate identity	Storytelling	Time-buyers	Standard phrases with simple meanings
	Turn claimers and holders		
	Personal turns of phrase		Fillers
Asserting group identity	"In" phrases		Turn-holders
	Group chants		Discourse shape markers
	Institutionalized forms of words		Repetitions of preceding input
	Ritual		
	Threats	Manipulation of information	Mnemonics
	Quotation		Lengthy texts one is required to learn
	Forms of address		
	Hedge		Rehearsal

There are three main pragmatic functions for social interactions. First, the manipulation of others involves meeting one's own needs from physical, emotional, or cognitive perspectives (e.g., "could you repeat?" "I wonder if…"). Second, asserting a separate identity means that the interactant either intends to be taken seriously (e.g., "you're never going to believe this, but…") or intends to be separated from others (e.g., "you know what I mean"). Third, in contrast, asserting group identity expresses overall membership (e.g., "dearly beloved") or places oneself in a hierarchy (e.g., "your highness"). There are also three main pragmatic functions for memory limitations. First, the function of the processing shortcut is to increase the speed or fluency of the utterances (e.g., "put the kettle on, will you?"). Second, time-buying phrases are used either for emphasis, rhythm, or fluency of the utterances (e.g., "draw a conclusion") or for buying more time by holding the floor in interactions (e.g., "let me just say…"). Third, the manipulation of information is mainly used to assist the interactant in memorizing the information discussed in interactions (e.g., "shall I compare…"). Wray and Perkins (2000) contend that both the pragmatic functions for social interactions and those for memory limitations represent two sides of the same coin. The pragmatic functions for social interactions ensure the smoothness of the interaction and the survival of all the interactants involved in one interaction. The pragmatic functions for memory limitations ensure that the production of utterances by an interactant is fluent.

2.3 Formulaic Language Use in Spoken English Discourse

Research on formulaic language in spoken English discourse has investigated both native English speakers and ESL learners (Ambele et

al., 2018; Bardovi-Harlig & Stringer, 2017; Erman & Lewis, 2022). Various data collection techniques have been utilized, including the oral discourse completion task, the retelling task, and the role-play task, resulting in distinct findings in each study. Ambele et al. (2018) discovered that ESL learners use formulaic language less frequently than native English speakers, whereas Erman and Lewis (2022) found that advanced L2 English speakers who had resided in an English-speaking environment used formulaic language more frequently than those at other English proficiency levels. The findings on formulaic language use in spoken English discourse differ from those on academic English writing, where ESL learners at a lower English proficiency level tend to use formulaic language more often than those at higher levels (Chen & Baker, 2016). The differing findings suggest that the use of formulaic language by ESL learners is linked to their English proficiency levels. Furthermore, investigations into the pragmatic functions of formulaic language employed by ESL learners have revealed that they may not have a complete understanding of the importance of formulaic language functions. In interlanguage communication, ESL learners use formulaic language to express certain functions influenced by their L1 background knowledge (Ambele et al., 2018; Bardovi-Harlig & Stringer, 2017).

In comparison to studies of formulaic language in written English discourse, research on formulaic language in spoken English discourse appears to be inadequate (Erman & Lewis, 2022). Previous studies have rarely utilized naturally occurring data, which may have limited the full exploration of formulaic language in spoken English discourse. Furthermore, the lack of systematic investigation into the use of formulaic language in spoken English discourse, unlike studies on written English discourse, has resulted in a deficient comparison of formulaic language usage among ESL learners at distinct English proficiency levels. Both of these limitations have contributed to an inadequate understanding of the use of formulaic language by ESL learners in interlanguage communication.

The present study investigates the frequencies and pragmatic functions of the formulaic language used by Thai ESL learners at different English proficiency levels in the interlanguage pragmatic field. It identifies the most frequently used formulaic language and the pragmatic functions in naturally occurring data. This study uses a corpus-driven approach (described in the next section), and the most frequent four-word lexical bundles are extracted for analysis. The study seeks to answer three research questions:

1) What is the most frequently used formulaic language by Thai ESL learners at the basic, intermediate, and advanced English proficiency levels in interlanguage communication?
2) What are the pragmatic functions of the formulaic language in spoken discourse used by Thai ESL learners at the basic, intermediate, and advanced English proficiency levels in interlanguage communication?
3) What are the main differences in the frequencies and the pragmatic functions of the formulaic language used by Thai ESL learners at the basic, intermediate, and advanced English proficiency levels in interlanguage communication?

3. Methodology

3.1 Participants and Data Collection

A total of 120 Thai ESL learners, comprising 64 males and 56 females, participated voluntarily in this study, with 40 participants at each English proficiency level. Participants were classified into the appropriate English proficiency level (Council of Europe, 2020) based on the score of an English test, such as TOEFL iBT. All participants were Thai nationals whose native language or L1 was Thai, and they had studied ESL for at least 12 years before data collection. They were all undergraduate students aged 18-23 years, from first year to fourth year, enrolled in eight faculties, such as the Faculty of Engineering and the Faculty of Science, at a public university in Bangkok, Thailand. Nine participants had travelled to English-speaking countries, such as the UK, with the longest stay being 10 days. However, no participant had lived in an English-speaking country for an extended period.

Each participant was paired with another participant at the same English proficiency level and asked to engage in a 20-minute English conversation in the genre of daily casual conversation. The aim of this study was to investigate the naturally occurring use of formulaic language by ESL learners. Each pair had complete freedom in selecting the topic of their conversation and could choose any place on the campus to have the conversation, such as the common study room at the library. The researcher did not appear during any of the conversations to ensure that participants felt comfortable and relaxed while talking with each other in English, and that the data collected were as natural as possible. Each conversation was recorded using an iPhone 11 Voice Memos, and the spoken data were transcribed into written form using CHAT conventions (MacWhinney, 2000). As a result, 20 samples of dyadic English conversations were collected from each of the 40 Thai ESL learners at each English proficiency level, resulting in a total of 60 samples of dyadic English conversations from the 120 Thai ESL learners.

3.2 Data Analysis

Learner corpus research (LCR) is a recent methodology employed in diverse ESL fields, including interlanguage pragmatics (Fernández & Staples, 2021; Staples & Fernández, 2019). LCR focuses on the naturally occurring interlanguage of language learners, and researchers may use a large machine-readable database of interlanguage to combine top-down and bottom-up analysis (Paquot & Granger, 2012). Previous research has shown that studies of formulaic language have been central to both corpus linguistics and LCR (Fernández & Staples, 2021; Paquot & Granger, 2012). The present study utilized LCR as the methodology, combining top-down analysis with Wray and Perkins' (2000) framework and bottom-up analysis with the corpus-driven approach.

This study adopted the use of four-word lexical bundles, both fixed and semi-fixed, which are deemed appropriate for studying formulaic language. Empirical evidence has demonstrated that four-word lexical bundles typically include three-word lexical bundles, while

five-word lexical bundles are less frequently encountered in discourse (Chen & Baker, 2016; Tan & Römer, 2022).

To extract the most frequently used four-word lexical bundles by Thai ESL learners, this study employed two frequency-based approaches. First, previous studies established frequency cut-off points, typically around 10 instances per million tokens, with an appearance in at least five different texts, which are considered standard for large corpora (Biber, 2009; Narkprom & Phoocharoensil, 2022). As the learner corpus in this study is relatively small compared to corpora with over one million tokens, the top 10 most frequently used four-word lexical bundles, utilized by at least five participants at the same English proficiency level, were selected as the cut-off point. The n-gram function in AntConc (Anthony, 2022) was used to retrieve the bundles. Three criteria were applied: first, bundles that were topic-dependent (e.g., the environment in Thailand) were excluded, as their recurrent use was solely due to being the topic of conversation (Chen & Baker, 2016). Second, overlapping bundles (e.g., "I don't know it" and "don't know it is") were manually examined by the researcher, as they may have originated from the same context. These were combined as five-word lexical bundles, for instance, "I don't know it is." Third, bundles without concrete semantic meaning (e.g., "and the the uhh") were excluded as they were considered fillers with no specific purpose between utterances. The mutual information (MI) calculated by AntConc (2022) was then used to determine a significant association of the co-occurrences found in the first stage. The purpose of examining the MI score is to determine the extent to which two linguistic tokens co-occur in each context. Paquot and Granger (2012) note that MI has seldom been used in studies of the formulaic language of ESL learners, partly due to the difficulty in identifying closely associated lexical bundles produced by ESL learners. Following previous research, an MI score above 3 is regarded as significant (Yeldham, 2020). The two stages of the frequency-based approach were employed to ensure that the four-word lexical bundles were indeed the most frequently used.

Furthermore, Wray and Perkins' (2000) framework was used to determine the pragmatic functions of each lexical bundle found in the corpus. Each pragmatic function was identified based on the concordancers provided by AntConc (2022). Differences in the frequencies and pragmatic functions used by Thai ESL learners at different English proficiency levels were identified based on the results of Research Questions 1 and 2.

4. Results and Discussion

4.1 Frequencies of Four-Word Lexical Bundles

For this study, a learner corpus was built entitled the Learner Corpus of Formulaic Language (LCFL). The corpus consisted of three sub-corpora: the Learner Corpus of Formulaic Language at the Basic Level (LCFLBL), the Learner Corpus of Formulaic Language at the Intermediate Level (LCFLIL), and the Learner Corpus of Formulaic Language at the Advanced Level (LCFLAL). LCFL contains 127,959 tokens: 26,158 tokens in LCFLBL, 40,562 in LCFLIL, and 61,239 in LCFLAL. Table 2 lists the seven most frequently used four-word lexical bundles in LCFLBL.

Table 2. Most frequently used formulaic language used by Thai ESL learners at basic English proficiency level

Formulaic language: Four-word lexical bundles	Frequencies	Number of interactants	MI score
** i don't know*	443	36	5.81
*i think it **	428	38	6.18
going to talk about	365	21	7.26
we are going to	314	24	5.88
what do you think	305	17	8.19
*i don't really **	285	13	5.45
*what i mean **	241	15	5.78

Table 2 shows that ESL learners at the basic English proficiency level used the semi-fixed four-word lexical bundle "* I don't know" the most in English conversations compared to other four-word lexical bundles, with 36 out of 40 participants (90%) using it. Six lexical items were found at the beginning of this lexical bundle used by ESL learners at the basic level: "emm", "uhh", "it", "I", "yes", and "no". The lexical items "emm" and "uhh" were used the most frequently (57.15%). Another three semi-fixed units were also found: "I think it *", "I don't really *", and "what I mean *". Four lexical items were found at the end of "I think it *": "is", "uhh", "was", and "it". Five were found at the end of "I don't really *": "uhh", "know", "understand", "go", and "do". Three were found at the end of "what I mean *": "is", "was", and "uhh". Moreover, three fixed four-word lexical bundles were found. Although "I think it *" was the second most frequently used four-word lexical bundle based on frequency, it was used by the highest number of participants (95%). The fixed unit "what do you think" had the highest MI score, indicating that this four-word lexical bundle is more strongly collocated than others. Table 3 shows the six most frequently used four-word lexical bundles in LCFLIL.

Table 3. Most frequently used formulaic language used by Thai ESL learners at intermediate English proficiency level

Formulaic language: Four-word lexical bundles	Frequencies	Number of interactants	MI score
i don't know it	314	31	3.78
** it's ok it's*	188	25	5.89
*i don't like **	94	18	3.55
*let's talk about **	68	18	5.97
*i think uhh **	65	29	4.10
** something like that*	64	22	4.05

The four-word lexical bundle "I don't know it" is the most frequently used formulaic language by ESL learners at the intermediate English

proficiency level, with 31 out of 40 participants (77.5%) using it. The other five lexical bundles are semi-fixed units. The second most frequently used one, "* it's ok it's", was used with "uhh", "emm", "it", "I", and "well" at the beginning. This finding may indicate the phenomenon of filler redundancy produced by ESL learners in naturally occurring data. A similar phenomenon was also found in the use of "I think uhh *" and " something like that". The filler "emm" was used most often at the end of "I think uhh *" (32 instances, 49.23%) and at the beginning of "* something like that" (27 instances, 42.19%). The lexical items at the end of "I don't like *" and "let's talk about *" mainly pertained to the topic under discussion (e.g., "I don't like people") and pronouns relating to the context (e.g., "let's talk about him"). Finally, this study found five four-word lexical bundles that met the requirements of both the top 10 frequencies and the MI score in LCFLAL, as illustrated in Table 4.

Table 4. Most frequently used formulaic language used by Thai ESL learners at advanced English proficiency level

Formulaic language: Four-word lexical bundles	Frequencies	Number of interactants	MI score
* *i think it*	411	26	4.12
what i mean is	395	21	4.41
i don't know *	185	19	3.31
* *something like that*	103	14	3.06
i mean it's just	74	15	3.55

Table 4 shows that two fixed four-word lexical bundles were found: "what I mean is" and "I mean it's just". The most frequently used four-word lexical bundle was a semi-fixed unit: "* I think it". Six lexical items were identified at the beginning of it: "I", "uhh", "it", "well", "emm", and "first". The two semi-fixed units, "I don't know *" and "* something like that", were also used by advanced-level ESL learners. Various lexical items were identified at the end of "I don't know ", such as "maybe" and "when", and at the beginning of "* something like that", such as "uhh".

Tables 2 to 4 present the most frequently used four-word lexical bundles by Thai ESL learners at each English proficiency level. The lexical bundle "I don't know" was identified as a common co-occurrence across all three English proficiency levels, indicating its frequent use among Thai ESL learners at different levels. This finding is consistent with prior research (Coane & Umanath, 2019; Hosoda & Aline, 2021) that reports frequent usage of "I don't know" by ESL learners from diverse linguistic backgrounds in English communication. Another commonly used co-occurrence found within the shared four-word lexical bundles across all three English proficiency levels is "I think". This observation is congruent with recent studies (Pan & Aroonmanakun, 2022; Wood, 2019) that report a high frequency of usage of "I think" by ESL learners. Pan and Aroonmanakun (2022) note that Thai ESL learners tend to "overuse 'I think'" (p. 198) in English communication. Additionally, the lexical bundles frequently used by Thai ESL learners in academic writing (e.g., "on the one hand"; Narkprom & Phoocharoensil, 2022) differ from those used by Thai ESL learners in interlanguage communication, indicating differences in the use of formulaic language by ESL learners with the same linguistic background across different English genres.

4.2 Pragmatic Functions of Four-Word Lexical Bundles at the Basic Level

Drawing on the framework proposed by Wray and Perkins (2000) and a corpus-driven approach, this study presents the pragmatic functions of the four-word lexical bundles that Thai ESL learners at the basic English proficiency level use most frequently in Table 5.

Table 5. Pragmatic functions of each four-word lexical bundle used by Thai ESL learners at the basic English proficiency level

Formulaic language: Four-word lexical bundles	Devices	Category of the pragmatic function	Pragmatic function
* *i don't know*	Devices for memory limitations	Time-buyers	Standard phrases with simple meanings
	Devices for social interactions	Asserting group identity	Hedge
i think it *	Devices for memory limitations	Time-buyers	Fillers
going to talk about	Devices for memory limitations	Processing short-cuts	Standard phrases
we are going to	Devices for memory limitations	Processing short-cuts	Standard phrases
what do you think	Devices for memory limitations	Processing short-cuts	Standard phrases
i don't really *	Devices for social interactions	Asserting separate identity	Turn claimers and holders
what i mean *	Devices for social interactions	Asserting separate identity	Personal turns of phrase

The findings indicate that basic-level ESL learners primarily employ four-word lexical bundles as devices to cope with memory limitations (five out of eight instances). The most commonly used lexical bundle, "*I don't know," serves two different pragmatic functions as a standard phrase to respond to the previous interactant, as illustrated in Excerpt 1.

B12: (.) how (.) did he do it ?

B11: &-emm (.) i don't know (.) no idea .

Excerpt 1

B11 uses the filler "emm" to begin the semi-fixed unit "*I don't know" in response to B12's question in Excerpt 1. Additionally, B11 adds

"no idea" to emphasize not knowing the answer. Prior research by Coane and Umanath (2019) has demonstrated that the use of "I don't know" reveals a state of "not having information in the knowledge base" (p. 152), indicating that the speaker lacks knowledge. However, the present study finds that "*I don't know" also functions as a hedge to mitigate potential tension in interactions, as illustrated in Excerpt 2.

B25: &-emm (..) but how do you know it ?

B26: &-uhh (.) the [/] the first one (.) it should (.) be right .

B25: (.) no (.) how [/] how right one ?

B26: &-uhh (..) i don’t know (.) maybe the middle (.) also ok .

Excerpt 2

In Excerpt 2, the two interlocutors are discussing how to play chess, and there is an argument between them. B26 uses "*I don't know" not only to respond to the previous interlocutor but also to mitigate the argument with B25, as indicated by the insertion of the filler "uhh" and the use of "maybe" thereafter. Previous research has found that Japanese ESL learners use "I don't know" to manage different types of epistemic stance, including hedging (Alharthi, 2015; Hosoda & Aline, 2021). This study further ascertains that ESL learners use "I don't know" as a hedging device in interlanguage communication.

Furthermore, the semi-fixed unit "I think it *" functions as a filler in interactions to buy time, as illustrated in Excerpt 3. The semantic sense of the utterance following "I think it *" is supposed to be that B09's parents took B09 to the place that the interlocutors are discussing at the moment. However, if "I think" is deleted from the original utterance, the semantic sense of the utterance remains the same. This finding supports previous research on "I think" (Ambele et al., 2018; Pan & Aroonmanakun, 2022), which found evidence of its pragmatic use as a filler by both native English speakers and ESL learners. The present study further reveals the use of "I think it *" as a filler between utterances to buy time in interlanguage communication.

B09: this (.) place &-uhh (.) yes it is .

B10: this is the place ?

B09: yes (.) and i think (.) it is place (.) when my parents take me many time .

Excerpt 3

Besides the first two lexical bundles, "going to talk about," "we are going to," and "what do you think" were used as standard phrases for the interlocutors to process the utterances more clearly and fluently. The lexical bundle "I don't really *" was used to assert the interlocutor's identity as a turn claimer. The last lexical bundle, "what I mean *," is discussed in the following section.

4.3 Pragmatic Functions of Four-Word Lexical Bundles at the Intermediate Level

Table 6 presents the pragmatic functions of the four-word lexical bundles that were most frequently used by Thai ESL learners at the intermediate English proficiency level.

Table 6. Pragmatic functions of each four-word lexical bundle used by Thai ESL learners at the intermediate English proficiency level

Formulaic language: Four-word lexical bundles	Devices	Category of the pragmatic function	Pragmatic function
i don’t know it	Devices for memory limitations	Time-buyers	Standard phrases with simple meanings
	Devices for social interactions	Asserting group identity	Hedge
** it’s ok it’s*	Devices for memory limitations	Time-buyers	Turn-holder
*i don’t like **	Devices for social interactions	Asserting group identity	Personal turns of phrase
*let’s talk about **	Devices for memory limitations	Processing short-cuts	Standard phrases
*i think uhh **	Devices for memory limitations	time-buyers	Filler
** something like that*	Devices for memory limitations	Processing short-cuts	Standard phrases

Similar to ESL learners at the basic level, those at the intermediate level employ "*I don’t know" either as a device for responding to a question to which no answer can be provided or as a device for mitigating potential tension. Additionally, "I think uhh *" is utilized as a filler between two utterances to buy more time. Furthermore, the standard phrase "let’s talk about *" is a conversational routine that directs the interaction towards ongoing topics. Moreover, Thai ESL learners use "I don’t like *" as a turn-taking device to share their opinions.

The semi-fixed unit "*it’s ok it’s" was frequently used as a turn-holder, allowing the interlocutor to continue the utterance, as illustrated in Excerpt 4.

I27: went (.) there before (.) i mean we .

I28: you and your sister (.) just lucky .

I27: &-uhh (.) not lucky (.) oh (.) yeah (.) ok (.) .

I28: &=laugh .

I28: well it’s ok it’s just (.) different life .

Excerpt 4

Lexical bundles centered on "OK" have been found to be frequently used with various pragmatic functions (Kiramba & Oloo, 2023). In Excerpt 4, the interlocutors are discussing "dream" places they had been to before. I27 had been to a place that I28 had never visited. After I28 expresses a wish to have visited the place too, I27 notices that the information about the place may lead to negative feelings for I28. Hence, I27 utters a series of spoken fillers ("oh," "yeah," "OK"), revealing I27's awkwardness at that moment. However, I28 uses a laugh to show that they were not serious about the topic. Subsequently, I28 utters "well, it's OK" to hold the floor, followed immediately by another utterance, "it's just different life," after which I28 attempts to use "*it's ok it's" to hold the floor for more utterances.

The semi-fixed unit "*something like that" was frequently used by the ESL learners at the intermediate level, as Excerpt 5 illustrates.

I05: this &-uhh (.) process can [/] can check (.) first way is right .

I05: this way is it right or wrong .

I05: and that (.) is it &-uhh (.) ok &-emm (.) something like that .

Excerpt 5

Thai ESL learners at the intermediate level tended to insert "*something like that" at the end of an utterance, especially after several exemplifications had been given (47 instances, 73.4%). This phenomenon was also found in Quammie-Wallen's (2021) study, in which "something like that" is marked as "vague language" in Hong Kong English (p. 13), indicating a lack of relevant knowledge in the context. The Thai ESL learners inserted "*something like that" when they were unable to provide more exemplifications; the use of "*something like that" seemed to make the end of the utterance more fluent. Moreover, another 17 instances (26.6%) were found in which the interlocutors inserted "*something like that" after some fragments in the utterances, such as the one in Excerpt 6.

I09: &-uhh (.) just go straight (.) and &-uhh (.) i remember (.) .

I09: &-uhh (.) long time (.) yeah (.) just (.) something like that .

Excerpt 6

The fragments in the utterances are generally attributable to the deficiency in the English-speaking skill of the ESL learners, resulting in the use of "*something like that" to make the end of the utterance more fluent.

4.4 Pragmatic Functions of Four-Word Lexical Bundles at the Advanced Level

Table 7 displays the pragmatic functions of the four-word lexical bundles that were most frequently used by Thai ESL learners at the advanced English proficiency level.

Table 7. Pragmatic functions of each four-word lexical bundle used by Thai ESL learners at the advanced English proficiency level

Formulaic language: Four-word lexical bundles	Devices	Category of the pragmatic function	Pragmatic function
* *i think it*	Devices for memory limitations	time-buyers	Filler
what i mean is	Devices for social interactions	Asserting separate identity	Personal turns of phrase
i don't know *	Devices for memory limitations	time-buyers	Filler
	Devices for social interactions	Asserting group identity	Hedge
	Devices for memory limitations	Time-buyers	Standard phrases with simple meanings
* *something like that*	Devices for memory limitations	Processing short-cuts	Standard phrases
i mean it's just	Devices for social interactions	Asserting separate identity	Personal turns of phrase

Consistent with previous findings, the semi-fixed units "*I think it" and "I don't know *" were frequently used by Thai ESL learners at the advanced level. The unit "*I think it" functioned as a filler, while the same pragmatic function was found in the use of "I don't know *", which was not found in LCFLBL and LCFLIL, as demonstrated in Excerpt 7.

A33: probably cause she's (.) jealous of you .

A34: what (.) ?

A34: no (.) what could i be jealous ?

A33: (.) i don't know (..) perhaps your everything ?

Excerpt 7

As illustrated in Excerpt 7, there is no tension in the context, which is also apparent in the recording where the tone of both interlocutors is relaxed. The use of "I don't know" by A33 may be interpreted as a standard phrase to indicate that A33 does not know the answer to the previous question, "what could I be jealous." However, the word "maybe" is uttered after "I don't know" with a longer pause. Using "perhaps," A33 provides a possible answer after the lexical bundle "I don't know" is uttered. Therefore, the semi-fixed unit "I don't know

*" is considered a filler for A33 to buy more time before the subsequent utterance.

Two lexical bundles, "what I mean is" and "I mean it's just," were used as personal turns. "I mean" was used to identify a personal explanation, interpretation, or opinion, as illustrated in the two excerpts below.

A07: it's dumplings from what (.) china ?

A08: no (.) it's actually not dumplings (.) it's like dumplings .

A08: &-uhh (.) it's with the dough outside right (.) and with the meat inside (.) it .

A07: what are you talking about ?

A08: &=laugh

A08: what i mean is this is not really the dumpling but it (.) has another name .

Excerpt 8

A15: i decide to study environmental because i maybe like it .

A16: you like it now still ?

A15: &-emm (.) yeah i (.) think so .

A15: i mean (.) it's just hard for me to [/] to some exams (.) but i still like it .

Excerpt 9

In both excerpts above, the interlocutors primarily used the lexical bundles to clarify their previous ambiguous utterances (i.e., A08 did not explain some type of food well; A15's utterance "yeah, I think so" did not fully answer the previous question "you like it now still?"). This further explanation, interpretation, or opinion-sharing marks the personal turn, as the interlocutors intended to identify their own perspectives. Pan and Aroonmanakun (2022) also found that Thai ESL learners frequently used "I mean" as a turn-taking device to illustrate personal explanation. This study revealed the use of lexical bundles centered on "I mean" by Thai ESL learners at different English proficiency levels in interlanguage communication.

4.5 Different Uses of Four-Word Lexical Bundles at Different English Proficiency Levels

Two similar four-word lexical bundles were identified at different English proficiency levels: one centered on "I don't know" and the other centered on "I think". Figure 1 and Figure 2 demonstrate the difference in the frequency and the number of participants who used these lexical bundles at different English levels.

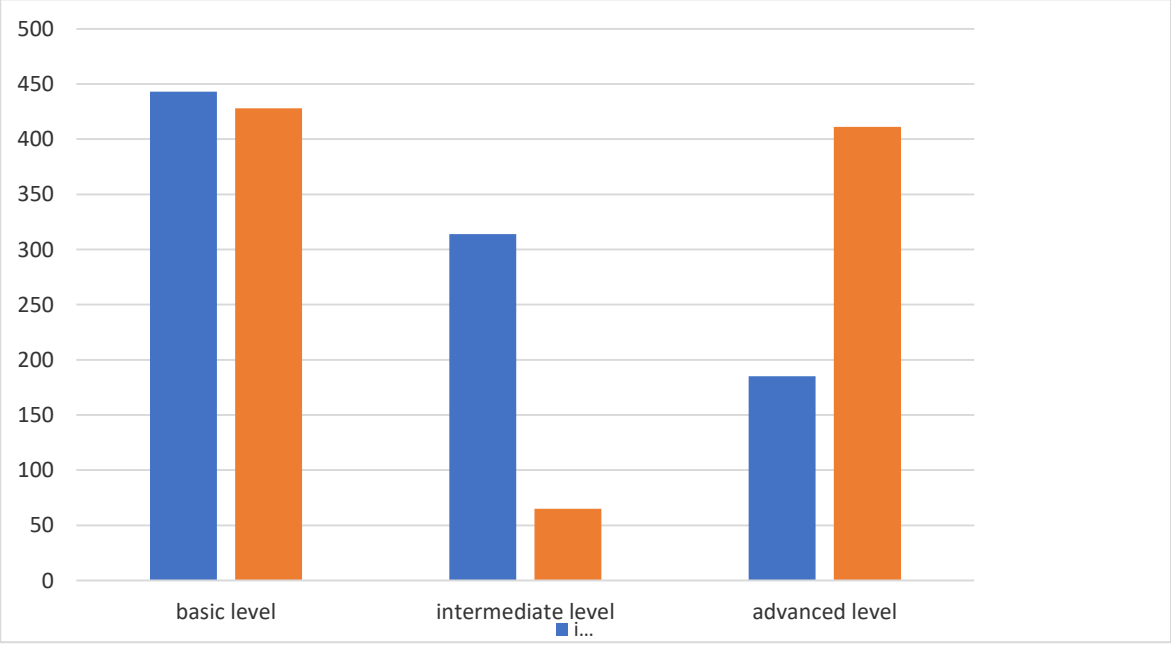

Figure 1. Frequencies of using the four-word lexical bundles centered by "I don't know" and "I think" by Thai ESL learners at three English levels

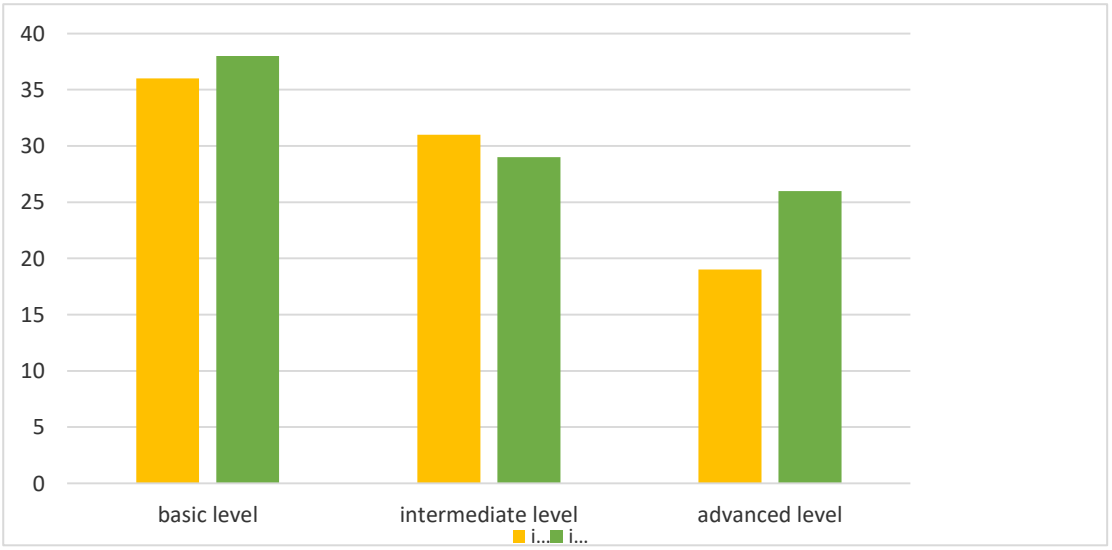

Figure 2. Number of participants who used the four-word lexical bundles centered by “I don’t know” and “I think” by Thai ESL learners at three English levels

As depicted in Figure 1 and Figure 2, both the frequency and the number of participants who used the lexical bundles centered on "I don’t know" decreased from the basic level to the advanced level. Concerning those centered on "I think", the number of participants who used it decreased from the basic level to the advanced level, and intermediate-level ESL learners utilized it much less frequently than those at the other two levels. These findings suggest that ESL learners at higher levels tend to use formulaic language centered on "I don’t know" less than those at lower levels, whereas ESL learners at the intermediate level do not rely on "I think" as much as those at the other two levels.

In addition to the two lexical bundles mentioned above, two other lexical bundles were frequently used at two different levels. One centered on "what I mean", which was found at the basic level, was used more often by advanced-level ESL learners in terms of both frequency and the number of participants. The lexical bundle "something like that", which was found at the intermediate and advanced levels, was likewise used more often as the English level increased. However, fewer participants at the advanced level used it than at the intermediate level, indicating an imbalance in its use by advanced-level ESL learners.

The different frequencies of use of similar lexical bundles indicate different uses of formulaic language by ESL learners with the same L1 backgrounds at different English proficiency levels. Furthermore, the formulaic language used by ESL learners at different English proficiency levels was distinct from each other, which is consistent with previous research.

Regarding the differences in the pragmatic functions of four-word lexical bundles, Figure 3 and Figure 4 depict the use of four-word lexical bundles as devices for social interaction and for memory limitations by Thai ESL learners at different English levels.

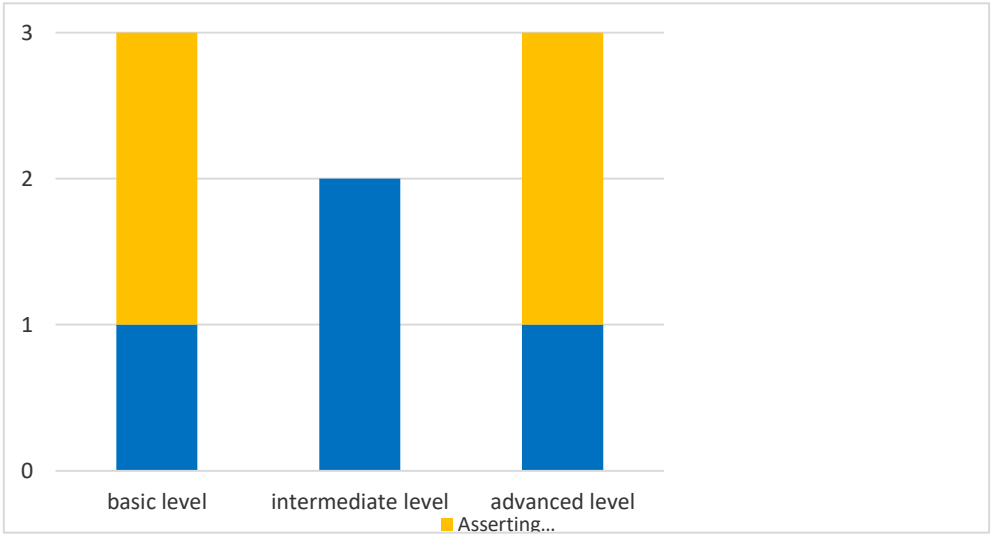

Figure 3. Proportion of using the four-word lexical bundles as devices for social interactions at three English levels

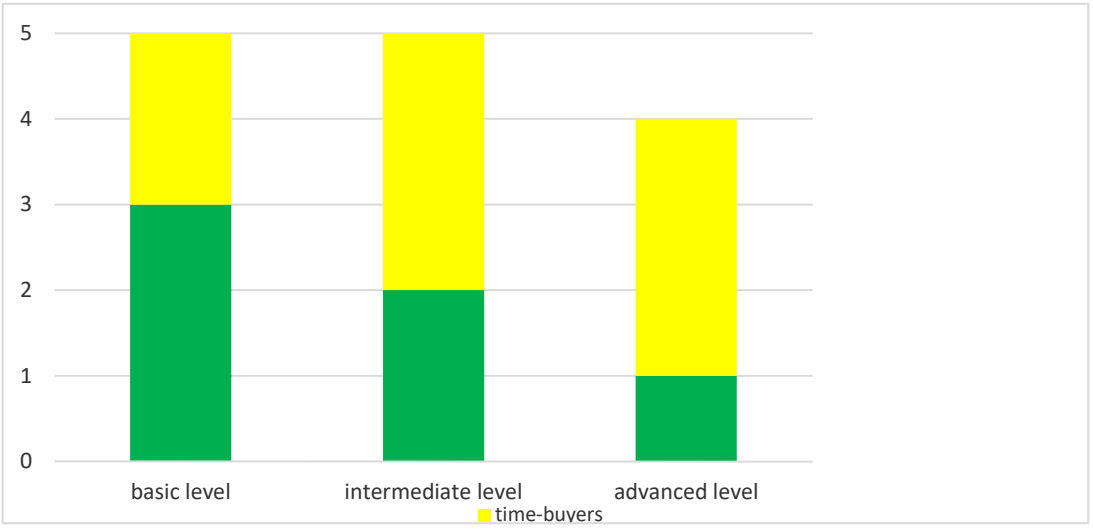

Figure 4. Proportion of using the four-word lexical bundles as devices for memory limitations at three English levels

As shown in Figures 3 and 4, intermediate-level Thai ESL learners did not use the four-word lexical bundles to assert a separate identity for social interactions. This phenomenon is a new finding since few studies have focused on the pragmatic functions of formulaic language used by intermediate-level ESL learners. The different pragmatic functions used by ESL learners at the three English proficiency levels indicate that the use of formulaic language is related to the English proficiency level. Moreover, the four-word lexical bundles used as memory shortcuts were more frequent at the basic level than at the advanced level, while the use of the four-word lexical bundles as time-buyers was more frequent among ESL learners at higher proficiency levels than among those at lower levels. Both findings may suggest that ESL learners at higher English levels tend not to rely on formulaic language to make their utterances fluent, unlike those at lower English levels.

5. Conclusion

The frequency lists reveal that individuals at higher levels of English proficiency utilize fewer types of four-word lexical bundles in interlanguage communication than those at lower levels. This finding indicates that ESL learners at higher levels demonstrate a greater level of flexibility in their use of English than those at lower levels. With regard to the pragmatic functions, learners at all three levels of English proficiency mainly use four-word lexical bundles as devices for social interactions to assert group identity, revealing that ESL learners tend to employ formulaic language for solidarity in interlanguage communication. Additionally, learners tend to utilize four-word lexical bundles as devices for memory limitations to buy more time between utterances. This finding indicates that ESL learners at different levels require certain formulaic language to accomplish utterances in interlanguage communication. Through these findings, it seems that ESL learners at various levels of English proficiency need formulaic language to complete the English communication. Thus, the input of formulaic language in second language acquisition is essential to ESL learners. Since ESL learners at lower levels reply on formulaic language more than the ones at higher levels, different formulaic language with various pragmatic functions should be guided to ESL learners at lower levels of English proficiency. Besides, instead of using fillers or redundant words, ESL learners may choose different formulaic language to buy more time in interlanguage communication, which helps the smoothness of the communication.

Regarding both frequencies and pragmatic functions, the outcomes imply that ESL learners at different English proficiency levels do not use four-word lexical bundles in the same manner in interlanguage communication. Moreover, the use of four-word lexical bundles in speech differs from their use in writing, as indicated by prior research results (Narkprom & Phoocharoensil, 2022). Both of these points suggest that future studies on formulaic language should consider the English proficiency level of the ESL learners and the language genre. This is because the formulaic language use by ESL learners at a certain level of English proficiency cannot represent all the ESL learners at other levels. The differences in the use of formulaic language in different language genres suggest that the formulaic language use relies on the language genre as an important factor that cause ESL learners to choose different types of formulaic language. Since few studies have concentrated on the use of formulaic language in spoken English discourse, future investigations should also examine formulaic language in different spoken genres employed by ESL learners with diverse L1 backgrounds to corroborate the findings of this research.

References

Alharthi, T. (2015). Adding more fuel to the fire: A study of attrition in formulaic sequences by adult learners. *Arab World English Journal, 6*(3), 230-243. https://doi.org/10.24093/awej/vol6no3.15

Ambele, E. A., Bonnsuk, Y., & Buddharat, C. (2018). Processing English formulaic expressions in situation-bound utterances: Strategies used by Francophone ESL learners in Thailand. *Arab World English Journal, 9*(3), 163-175. https://dx.doi.org/10.24093/awej/vol9no3.11

Anthony, L. (2022). *AntConc* (MacOS 10/11, 4.1.3 Version) [Computer Software]. Retrieved from

https://www.laurenceanthony.net/software/antconc/

Bardovi-Harlig, K., & Stringer, D. (2017). Unconventional expressions: Productive syntax in the L2 acquisition of formulaic language. *Second Language Research, 33*(1), 61-90. https://doi.org/10.1177/0267658316641725

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics, 14*(3), 275-311. https://doi.org/10.1075/ijcl.14.3.08bib

Chen, Y. H., & Baker, P. (2016). Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics, 37*(6), 849-880.

Coane, J. H., & Umanath, S. (2019). I don't remember vs. I don't know: Phenomenological states associated with retrieval failures. *Journal of Memory and Language, 107*, 152-168. https://doi.org/10.1016/j.jml.2019.05.002

Council of Europe. (2020). Common European Framework of Reference for Languages: Learning, teaching, assessment-Companion volume. Council of Europe Publishing. Retrieved from www.coe.int/lang-cefr

Ebrahimi, F., Namaziandost, E., Ziafar, M., & Seraj, P. M. I. (2021). The effect of teaching formulaic expressions through contrastive lexical approach on Iranian pre-intermediate EFL learners' writing skill. *Journal of Psycholinguistic Research, 50,* 1087-1105. https://doi.org/10.1007/s10936-021-09778-z

Erman, B., & Lewis, M. (2022). Formulaic language in L1 and advanced L2 English speech: multiword structures in the speech of two Swedish groups compared to a group of L1 English speakers. *Text & Talk (Online),* 1-22. https://doi.org/10.1515/text-2021-0090

Erman, B., & Warren, B. (2000). The idiom principle and the open-choice principle. *Text & Talk, 20,* 29-62. https://doi.org/10.1515/text.1.2000.20.1.29

Fernández, J., & Staples, S. (2021). Pragmatic approach. In N. Tracy-Ventura & M. Paquot (Eds.). *Routledge handbook of second language acquisition and corpora* (pp. 242-251). Taylor & Francis Group.

Hosoda, Y., & Aline, D. (2021). Deployment of *I don't know* and *wakannai* in second language classroom peer discussions. *Text & Talk, 42*(1), 27-49. https://doi.org/10.1515/text-2019-0275

Hyland, K. (2008). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18*(1), 41-62. https://doi.org/10.1111/j.1473-4192.2008.00178.x

Kiramba, L. K., & Oloo, J. A. (2023). "It's OK. She Doesn't Even Speak English": Narratives of language, culture, and identity negotiation by immigrant high school students. *Urban Education, 58*(3), 398-426. https://doi.org/10.1177/0042085919873696

Leelasetakul, M. (2019). *Lexical bundles use by Thai undergraduate learners of English.* [Unpublished doctoral dissertation]. Pennsylvania State University.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Lawrence Erlbaum Associates.

Narkprom, N., & Phoocharoensil, S. (2022). Lexical bundles in native English speakers' and Thai writers' dissertations. *GEMA Online Journal of language Studies, 22*(3), 43-62. http://doi.org/10.17576/gema-2022-2203-03

Ogawa, C. (2021). Revised 4/3/2 task: Fluency training with formulaic language in the EFL classroom. *The Journal of Asia TEFL, 18*(4), 1108-1127. http://dx.doi.org/10.18823/asiatefl.2021.18.4.3.1108

Pan, Z., & Aroonmanakun, W. (2022). A corpus-based study on the use of spoken discourse markers by Thai EFL learners. *LEARN Journal: Language Education and Acquisition Research Network, 15*(2), 187-213.

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics, 32,* 130-149. http://doi.org/10.1017/S0267190512000098

Quammie-Wallen, P. (2021). Vague language in Hong Kong English, 'Something like that': A comparative corpus investigation into a defining feature of English in Hong Kong. *English Today, 37*(1), 13-25. https://doi.org/10.1017/S0266078419000415

Rajeswaran, C. M. (2019). Task, formulaic language, and role play for developing ESL students' academic language. *Indonesian Journal of Applied Linguistics, 8*(3), 536-544. https://doi.org/10.17509/ijal.v8i3.15252

Staples, S., & Fernández, J. (2019). Corpus linguistics approaches to L2 pragmatics research. In N. Taguchi (Ed.), *The Routledge handbook of second language acquisition and pragmatics* (pp. 241-254). Routledge. https://doi.org/10.4324/9781351164085-16

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes, 12,* 214-225. https://doi.org/10.1016/j.jeap.2013.05.002

Tan, Y., & Römer, U. (2022). Using phrase-frames to trace the language development of L1 Chinese learners of English. *System, 108,* 1-10. https://doi.org/10.1016/j.system.2022.102844

Wang, J., & Halenko, N. (2022). Developing the use of formulaic language for study abroad: a targeted instructional intervention. *The Language Learning Journal, 50*(4), 409-426. https://doi.org/10.1080/09571736.2022.2088446

Wood, D. (2019). *Classifying and identifying formulaic language*. Routledge. https://doi.org/10.4324/9780429291586-3

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press. https://doi.org/10.1017/CBO9780511519772

Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford University Press.

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language & Communication, 20*, 1-28. https://doi.org/10.1016/S0271-5309(99)00015-4

Yeldham, M. (2020). Does the presence of formulaic language help or hinder second language listeners' lower-level processing? *Language Teaching Research, 24*(3), 338-363. https://doi.org/10.1177/1362168818787828

Appendix

Transcription Conventions

&-	Marking filled pauses
(.)	Unusual short pause
(..)	Unusual middle pause
[/]	Single word repetition
<>[/]	Multiple words repetition
[//]	Retracing
xxx	Unintelligible utterance
&=laughs	Speaker's laugh