Chabahar Maritime University
Iranian Journal of English for Academic Purposes ISSN: 2476-3187
IJEAP, 2023, 12(3), 56-72 (Previously Published under the Title: Maritime English Journal)

Formulaic Sequences in Learners' Spoken English: A Comparative Corpus-based Study between Native and Non-Native Speakers of English

[1]Ashraf Vaziri
[2]Hamed Barjesteh*
[3]Atefeh Nasrollahi Mouziraji

Research Paper IJEAP-2307-1977 DOR: 20.1001.1.24763187.2023.12.3.4.6
Received: 2023-07-06 Accepted: 2023-09-25 Published: 2023-09-30

Abstract: Formulaic sequences are commonly employed in general and educational communication, and come up with the construction of rationality in talking and non-verbal language, in addition to playing an essential role in the conception of communication. The current study examined the rate of formulaic sequences among competent non-native university students, structurally and functionally. Notably, a comparative corpus-based study was conducted to probe the accomplishment of formulaic sequences through administering the conversation in group discussions between native and Iranian speakers of English. For this purpose, four-word formulaic sequences were drawn out from a corpus of 21 group discussions and were categorized following Biber et al. (2004) taxonomies. Michigan Corpus of Academic Spoken English (MICASE) was employed as a native corpus. The results uncovered that native speakers employed more formulaic sequences than EFL peers. Besides natives utilized 'discourse organizing bundles' in functional classification, and 'noun and prepositional phrases' in structural classification. However, non-native speakers used 'stance expressions' in functional classification and 'verb phrase fragments' in structural classification. The outcomes present various educational suggestions for EFL teachers and learners.

Keywords: Corpus Analysis, Formulaic Sequences, Learner Corpora, Lexical Bundles, Speaking Skill

Introduction

The requisite for speaking dexterity in English has been severely increasing in consequence of the fortifying social location of EFL in support of making universal connection. Speaking is the most challenging skill for EFL learners. EFL learners have problem in speaking confidently though they expend a lot of time and funds on learning English in various language departments but still discern speaking difficult to grasp (Mohammadi & Enayati, 2018). As Yang (2020) believed, in order to overcome speaking problems and attain native-like fluency and accuracy, it is better for language learners to acquire language chunks and formulaic language. Thomson et al. (2023) explained that multi-word phrase represents formulaic sequences beyond the word level that are frequent, recognizable and expected.

[1] PhD Student of TEFL, AshrafV294@Yahoo.com; Department of English Language and Literature, Ayatollah Amoli Branch, Islamic Azad University, Amol, Iran.

[2] Associate Professor of Applied Linguistics (Corresponding Author), ha_bar77@yahoo.com; Department of English Language and Literature, Ayatollah Amoli Branch, Islamic Azad University, Amol, Iran.

[3] Assistant Professor of TEFL, nasrollahi.atefeh@gmail.com; Department of English Language and Literature, Ayatollah Amoli Branch, Islamic Azad University, Amol, Iran.

As multi-word units in language learning play crucial position in spoken language, the center of attention of the majority of researches was on the significance of these expressions in constructing fluency and confidence (Pawley & Syder, 1983; Saadatara et al., 2023). Besides, it is stated that studying such lexis provides some familiarity for learners, while they are accumulated in mind and recovered in a while (Wray & Perkins, 2000). The individual cognitive system accumulation and a procedure of data consisted of language, proficiently by employing prefabricated multi-word compositions (Lee, 2020). The significance of applying these expressions includes two considerable purposes; the eloquent role and the capability to provide comprehensible communication (Wang & Kaatari, 2021). Because of this, Carrol and Coklin (2020) proved that studying word chunks and utilizing properly is imperative in language acquirement.

Wray and Perkins (2000) explained formulaic sequences like the clusters of words which come out respectively or occasionally in content. In other words, formulaic language may be described as a vast string of words. Compared to formulaic language structures, lexical bundles are extensively considered. On the other hand, Biber et al. (2004) and Lee (2020) declared, bundles are common and assessable and also Wood (2015) stated that bundles' incidence is enveloping in group discussions because of their essential position. As some researchers (Cortes, 2004; Hyland, 2012; Xu & Wijitsopon, 2023) defined, 'lexical bundles' are the compositions of more than two terms that are continually utilized within discussion related to the context. Wood's (2015) explanation of 'lexical bundles' was somehow the same as mentioned researches; he described them as the compositions of terms which are three or more terms, recognized in the context applying particular software regarding occurrence and string norm.

'Structural and functional' classifications of 'lexical bundles' are mostly focused. A purposeful categorization was produced by Biber et al. (2004) to make sure that 'lexical bundles' are main part of speech according to university collected data, which proposed three significant functional classifications: 'stance bundles, discourse organizers, and referential phrases'. The outcome of their research indicated that while the majority of formulaic series link two constitutions of two clauses, having solid syntactic connection and can be recognized as three primary structural forms: 'verb phrases fragments', 'clause fragments', and 'noun and preposition phrase fragments'. Some researchers (Oktavianti & Sarage, 2021) considered the structure, some others like Budiwiyanto and Suhardijanto (2020) and Wachidah et al. (2020) focused on functional sets. While Biber et al. (2004) revealed that functional structures differ through various verbal and non-verbal contexts. In current years, it has seen a rising concentration on comparative investigations into bundles. Shirazizadeh and Amirfazlian (2021) centralized on written context in order to investigate various features of 'lexical bundles'. Crisianita and Mandasari (2022) investigated 'lexical bundles' in speaking context, Yang (2017) considered them in text genres, and Kwary et al. (2017) studied bundles in some courses of study. As well, different professional writers may have special utilization attitudes to lexical bundles and considered them in a various method (Fajri et al., 2020).

Literature Review

A large number of investigations regarding lexical bundles have concentrated on non-verbal language through the goal of examining various aspects of lexical bundles in non-verbal resources among different orders (Hyland, 2008; Oktavianti & Prayogi, 2022; Shin, 2018; Strunkyt & Jurkūnait, 2008; Wingate, 2012). Considering lexical bundles in argumentative essays, Pan et al. (2016) and some researchers came to the same conclusion which NN students employed 'stance bundles' more than other sub-categories and they have some misuses of bundles (Bychkovska & Lee, 2017; Salazar, 2014). Regarding employing bundles in theses, Sugiarti et al. (2018) detected that students apply 'research-based bundles' and 'text-based bundles' more. Yakut et al. (2021) also found the same result in considering N and NN authors. In analyzing essay writings of Chinese students, Yang and Fang (2021) explored that 'research-based

bundles' are the greatest commonly applied. Investigating bundles in three different types (such as: textbooks, theses, articles) represented variant outcomes, 'prepositional bundles' in textbooks and 'noun bundles' in theses and articles were employed more (Shirazizadeh & Amirfazlian, 2021). Xuan and Kim (2022) considered lexical bundles in "English Education" and "Linguistics" dissertations and proved that, in structural category, prepositional phrases were the largest proportion in two disciplines and functionally, research-oriented bundles were the biggest frequency.

Hyland (as cited in Zipagan & Lee, 2018) declared that, still with the noteworthiness of the lexical bundles in increasing communicative competence and in evaluating learners' progress (Staples et al., 2013), few studies have considered using bundles in speaking English. Examining the phraseology of American mainstream film scripts, Xu and Wijitsopon (2023) found spoken formulaic expression and descriptive expressions, like place-referential and action-related lexical bundles. The discoveries of Hadizadeh and Vefali (2022) investigation showed that advanced learners employed several lexical bundles with changeable alternations and structures which displayed learners' socialisation in communication. Cancino and Iturrieta (2022) studied the effect of lexical bundles on speaking proficiency and concluded that those learners that were exposed by lexical bundles outperformed better than the others. Wang and Kaatari (2021) studied the aspects of bundles in spoken educational context and considered the verb 'say' in the corpus. Tavakoli and Uchihara (2020) studied the correlation between having fluency in speaking and employing of formulaic strings among different dexterity rates and found that advanced learners use sequences more competently comparing to lower levels. Checking stance markers in EFL speakers with different first language revealed that learners are more proficient to choose appropriate markers in their talking (Gablasova et al., 2017). Kwon and Lee (2014) found that non-native English teachers employed stance expressions in their speaking more than the other categories. In Indonesia, Maulany (2013) performed an investigation to specify the impacts of formulaic sequences on speaking ability in a primary school. Throughout these tests, Maulany checked five features of speaking ability. Analyzing outcomes of interview proved that the learners' scores became better and their understanding also improved.

Despite the significance and recurrent application of four-word bundles, few studies have been done about their specific use in group discussions. In Iran, lexical bundle has not been extensively considered, so the current research intended to fill up the gap by recognizing and scrutinizing common four-word lexical bundles 'structurally and functionally' within speaking group discussions. To tackle the problems of this research, tried to provide answers to the following questions:

Research Question One: Which lexical bundles are used generally in speaking of undergraduate university students' group discussions?

Research Question Two: What are the 'structural and functional' characteristics of extracted bundles?

Research Question Three: What are the resemblances and distinctions between two corpora, considering bundles' 'occurrence', 'structure' and 'function'?

Methodology

Iranian Corpus

Iranian EFL learners' spoken corpus of this study has been extracted from 21 group discussion transcriptions containing various issues such as: *Saving money, Is it good? Does it have any advantages or disadvantages? ; Exploring the past: What is your idea? ; Life begins at 40*. The topics of discussions were chosen from the Alexander (1970) "For and Against" book. The participants of the group discussions were 18 female and 10 male undergraduate university students (divided in 7 group

discussions) with age range of 20-26 from Universities of Khaje Nasir and Elm o Sanaat. All participants selected for the study were involved in English learning at least for 6 years. The participants were designated on the basis of convenience sampling, so they differed in their age, gender, and years of learning experience. Each group discussions had 4 members and all groups discussed 3 topics during the time ranging about 15 to 20 minutes.

Native Corpus

Between the years of 1997- 2001, the "University of Michigan's English language Institute" collected the "Michigan Corpus of Academic Spoken English" (MICASE), accompanied by 200 full hours in an educational speech which was taped as well as transcribed (Simpson-Vlach & Leicher, 2006). By the number of 152 speech occasions, they proved that MICASE comprises four academic partitions, such as: "Physical Sciences and Engineering", "Biological and Health Sciences", "Humanities and Art", and "Social Sciences and Education". Corpus of "Humanities and Art", and "Social Sciences and Education" were utilized in this research.

Corpus Tool

AntConc Software

To analyze the transcripts of group discussions and extract lexical bundles, the researchers need an appropriate tool which facilitates analysis. AntConc 3.3.0 was used as the primary analysis software in this research (Anthony, 2019). This software presents its users with the following analysis instruments: "*concordance, files view, clusters, N-grams, collocates, word and keyword list"*.

Taxonomy

The current research utilized Biber et al. (2004) taxonomies of function and structure, as it is one of the most practical taxonomies for analyzing spoken data. Both classifications are illustrated in the subsequent tables.

Table 1

Taxonomy of Lexical Bundles' Function (Biber et al., 2004, p. 384)

Set	Sub-set
Stance expressions	epistemic stance
	personal
	impersonal
	attitudinal/modality stance
	desire
	personal
	obligation/directive
	personal
	impersonal
	intention/prediction
	personal
	impersonal
	ability
	personal
	impersonal

Discourse organizers	topic introduction/focus topic elaboration/clarification
Referential expressions	identification/focus imprecision specification of attributes quantity tangible framing attributes intangible framing attributes time/place/text reference place reference time reference text deixis multi-functional reference

Table 2

Taxonomy of Lexical Bundles' Structure (Biber et al., 2004, p. 381)

1. Lexical bundles that incorporate VP
1a. (connector +) 1st/2nd person pronoun + VP
1b. (connector +) 3rd person pronoun + VP
1c. discourse marker + VP
1d. VP (with non-passive Verb)
1e. VP with passive verb
1f. yes-no Q
1g. wh-Q
2. Lexical bundles that incorporate dependent clause
2a. 1st/2nd person pronoun + dependent clause
2b. wh-clause
2c. if-clause
2d. (verb/adjective +) to-clause
2e. that-clause
3. Lexical bundles that incorporate NP/PP
3a. (connector +) noun phrase with of-phrase
3b. NP with other post-modifier
3c. other NP expressions
3d. PP expressions
3e. comparative expressions

Procedure

Research data have been collected with the group discussions accomplished by Iranian undergraduate learners, regarding definite subjects determined for them by the researcher. Their discussions were primary recorded and have been kept.

Since MICASE corpus was used as the source for this research, researcher transcribed Iranian's talking in their groups with the same method which was conducted in MICASE. For operating AntConc in analyzing data, Notepad format is necessitated, so all discussions have been typed and saved through text format using a computer program. All common four-word lexical bundles in every sub-set were recognized. The rate of recurrence cut-off for the current study was determined four in the corpus; it is comparable to 12 incidences for each million terms. It is in line with Jablonkai's (2009) study.

Throughout transcription, nothing has been changed and even participants' errors have not been modified. Afterward, the extracted lexical bundles were categorized based on the classifications which were mentioned and explained before. In order to do the procedure of classification more precisely, the transcripts were verified by researcher's colleague, and modifies were done if needed.

Results

Distribution of the Target Bundles

A first review of the two lists, native speakers (NSs) and non-native speakers (NNSs) of extracted bundles disclosed that native speakers' discussions comprised more bundles compared to NNS. By and large, an entire 505 target clusters were detected in native corpus and 315 bundles in Iranian English speakers' corpus. The results showed that 67 varied bundles were in native and 45 in NNSs corpus. The issue that native speakers utilized more lexical bundles than EFL speakers has been proved by (Adel & Erman, 2012; Karabacak & Qin, 2013; Kashiha & Chan, 2015) in previous researches.

Table 3

Division of Lexical Bundles in N and NN Corpora

Groups	Entire No. of N-grams types	Entire No. of N-grams token
Native Speakers	67	505
Non- native Speakers	45	315

Structural Division of Lexical Bundles

As well as the occurrence table, disintegration of the corpus demonstrated that undergraduate learners employed a mixture of constructions of lexical bundles in their talking. Analyzing results explained that nearly all the target bundles comprised 'phrasal' by contrast 'clausal' type. Table 4 provides the number and percent of the most important syntactic forms of the bundles within discussions. It is displayed, native speakers used 'noun and prepositional phrase' (50.74%) more than 'verb phrase' (34.33%) and 'dependent clause' (14.92%). Whereas, non-native speakers showed more tendency to verb phrases (46.67%), and then to noun and prepositional phrase fragments (40%). It is interesting to state, native speakers, as well as non-native groups, had more or less similar use of 'dependent clause', native speakers (14.92%) and non-native speakers (13.33%).

Table 4

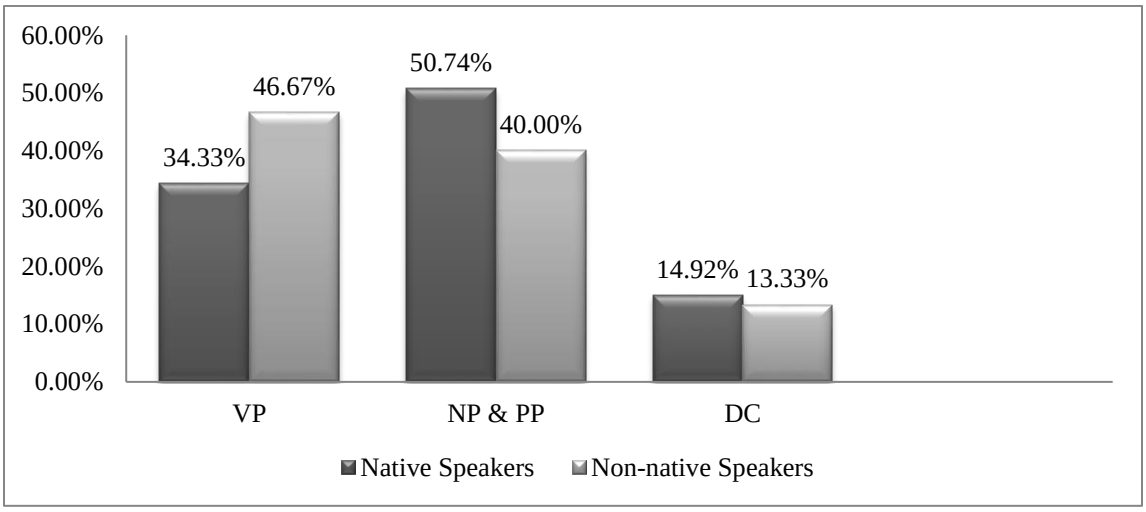
Structural Division of Lexical Bundles in N and NN Corpora

Structural Types	N No. (%)	NN No. (%)
Verb phrase	23 (34.33)	21 (46.67)
Noun phrase and prepositional phrase	34 (50.74)	18 (40)
Dependent clause	10 (14.92)	6 (13.33)
Total	67	45

Figure 1 compares various structural classifications of lexical bundles within both corpora.

Figure 1

A Distinction of Different Structural Classifications of Bundles in both N and NN Corpora



Verb Phrase Fragments

Considering 'verb phrase', results showed that Iranian learners utilized 21 'verb phrase' and native speakers employed 23 bundles. It is concluded that Iranian learners preferred 'verb phrase' more than the two other categories to express their opinion (*I want to talk*), showing their consensus or the opposite (*I agree with you*), request for further details or establish an issue by providing 'Wh-questions' (*what is your idea*), or emphasizing the problem by applying passive structure (*is on the basis*).

Table 5

VP in N and NN Corpora

Structural Sets	Sub-sets	N No. (%)	NN No. (%)	Sample
1. Lexical bundles that incorporate VP	a. 1st/2nd person pronoun + VP	3(4.48)	4 (8.88)	I don't want to
	b. 3rd person pronoun + VP	5 (7.46)	3 (6.66)	This is one of
	c. discourse marker + VP	3 (4.48)	4 (8.88)	I think this is

	d. verb phrase (with non-passive verb)	4 (5.97)	3 (6.66)	Let's look at the
	e. verb phrase (with passive verb)	2 (2.98)	1(2.22)	Has been proved that
	f. yes/ no question	3 (4.48)	2 (4.45)	Do you agree with
	g. wh-question	3 (4.48)	4 (8.88)	What do you mean
Total		23 (34.33)	21 (46.67)	

Noun and Prepositional Phrase

Besides, 'noun and prepositional phrase' counted for approximately 50.74% of the whole quantity of lexical bundles in N corpus and 40% in NN corpus. As the researcher stated in the following table, the use of 'prepositional phrase expressions' with 17.91% are on the top of the sub-categories and after that NSs used 'noun phrase with of- phrase' with 9 individual bundles (or 13.43%). While for non-native corpus, 'noun phrase with of-phrase' with 6 different bundles (13.33%) is the most frequent sub-category. After that, PP expressions and other 'noun phrase' wordings were used with equal numbers 5 (11.11). Based on the results that were proved in this research, using comparative expressions is the last sub-category that has been figured on by the NSs with (2.98%) and NNSs have not used them at all.

Table 6

NP and PP in N and NN Corpora

Structural Sets	Sub-sets	N No. (%)	NN No. (%)	Sample
2. Lexical bundles that incorporate NP/PP	a. NP with of-phrase	9 (13.43)	6 (13.33)	One of the most
	b. NP with other post-modifier	4 (5.97)	2 (4.45)	The way in which
	c. other NP expressions	7 (10.44)	5 (11.11)	Thank you very much
	d. PP expressions	12 (17.91)	5 (11.11)	At the same time
	e. comparative expressions	2 (2.98)	-	As well as the
Total		34 (50.74)	18 (40)	

Dependent Clause

As table 7 indicates, native speakers used 10 (14.92%) dependent clauses. The most prevalent sub-set within two corpora is item 'a' (*for example: I don't figure out*). For native corpus it is 5.97% and for non-native corpus 8.88%. The second sub-set is 'Wh-clause fragment', for native corpus is 4.48% and for non-native corpus is 2.22%. It is important to say that neither NS nor Iranian learners employed 'that clause' in their discussions. As it is reported in Kashiha and Chan (2015) research, since some structures are employed less, speaking schedules should be developed by utilizing lexical bundles.

Table 7

Dependent Clause in N and NN Corpora

Structural Sets	Sub-sets	N No. (%)	NN No. (%)	Sample
3. Lexical bundles that	a. 1st/2nd person pronoun +dependent clause	4 (5.97)	4 (8.88)	I don't know whether

incorporate	b. wh-clause	3 (4.48)	1(2.22)	When you say that
dependent clause	c. if-clause	2 (2.98)	1(2.22)	If you think about
	d. (verb/adjective)+ to-clause	1 (1.49)	-	To think how to
	e. that clause	-	-	Is that there is
Total		10 (14.92)	6 (13.33)	

Functional Division of Lexical Bundles

Findings show that spoken language in group discussions in the native corpus is dominated by 27 discourse organizers (40.29%), followed by 21 stance expressions (31.34%), and 19 referential expressions (28.35%) respectively. As concluded from Jukneviciene's (2009) investigation, 'stance expressions' are utilized in spoken context, while the two other classifications of functional bundles are employed in written context more (O'Keeffe et al., 2007). It can be realized that Iranian English speakers employed 42.22% of 'stance expressions' and native 31.34%. On the contrary, native speakers demonstrated more tendencies to apply 40.29% of 'discourse organizers', which was 37.77% in Iranian corpus. The N and NN corpora had related proportions in the employment of 'referential expressions' (28.35% and 20% in order).

Table 8

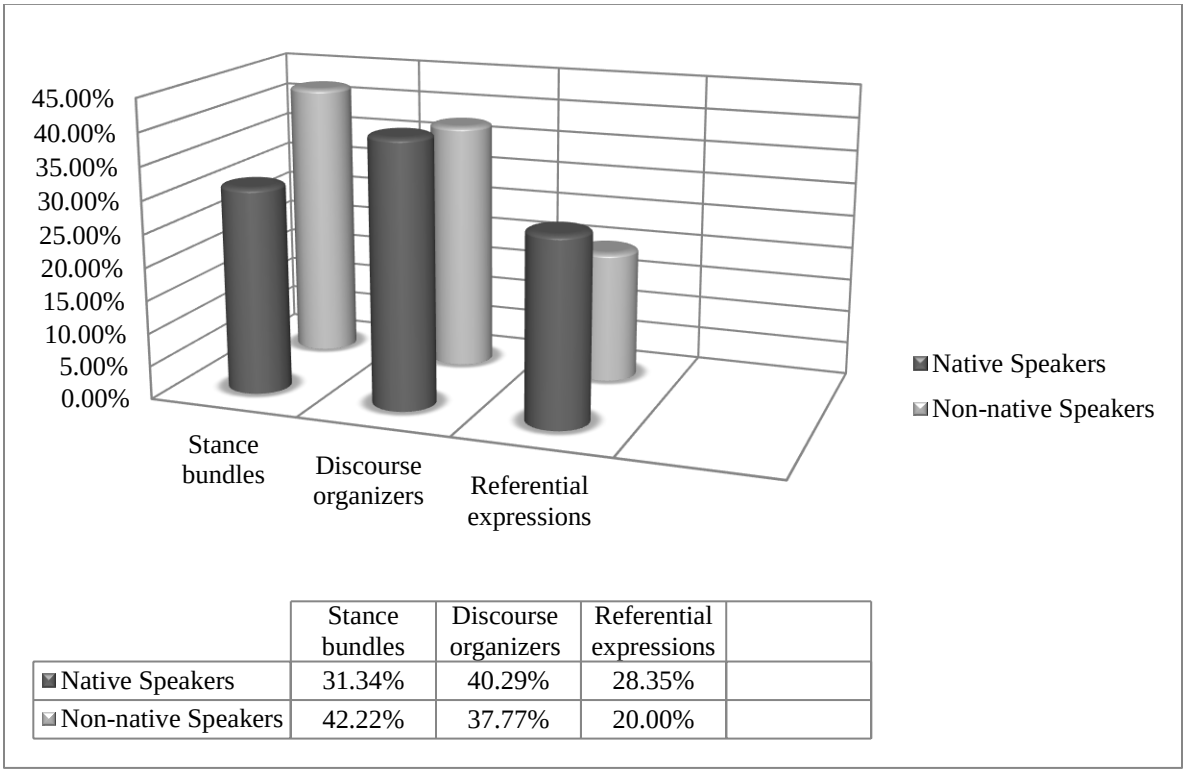
Functional Division of Lexical Bundles in N and NN Corpora

Sets	N No. (%)	NN No. (%)
Stance expressions	21 (31.34)	19 (42.22)
Discourse organizers	27 (40.29)	17 (37.77)
Referential expressions	19 (28.35)	9 (20)
Total	67	45

Figure 2 compares the different functional sets of lexical bundles within N and NN corpora.

Figure 2

A Distinction of Different Functional Categories of Bundles in both N and NN Corpora



	Stance bundles	Discourse organizers	Referential expressions	
Native Speakers	31.34%	40.29%	28.35%	
Non-native Speakers	42.22%	37.77%	20.00%	

Stance Bundles

Table 9 revealed that 'stance bundles' (42.22%) were the first sub-set that was utilized by Iranian learners. The researcher came to the conclusion that Iranian English speakers' propensity to carry out their discussion is using 'stance bundles' as an important function. This result is in contrast with native speakers' tendency. But in this category, both native and Iranian learners utilized 'epistemic stance expressions' more than other sub-sets, (11.94% and 15.55% respectively). One phrase that is used by EFL learners more is (*I don't think that*) which showed the speaker's uncertainty. Native speakers also used different forms of the modals like: 'would', meanwhile in this research, Iranian learners utilized the verb 'think'.

Considering 'stance expressions', Intention/Prediction was more common in both native (7.46%) and non-native (13.33%) corpus. But, there is variation in the structure of bundles utilized. Non-native speakers utilized phrases, like: (*I am going to*), to demonstrate the plan. Native speakers employed complicated utterances, like: (*I was supposed to*). There weren't many instances in B2 and B4 sub-sets in two corpora. For the B4 sub-set, native speakers used (4.48%) and non-native (6.67%).

Table 9

Stance Bundles in N and NN Corpora

Functional Sets	Sub-sets	N No. (%)	NN No. (%)	Sample
1. Stance bundles	A. epistemic stance	8 (11.94)	7 (15.55)	I think it was
	B. attitudinal/modality stance			
	B1.desire	2 (2.98)	1 (2.23)	I don't want to

	B2. obligation/ directive	3 (4.48)	2 (4.45)	Do you have to
	B3. intention/prediction	5 (7.46)	6 (13.33)	I am going to
	B4. ability	3 (4.48)	3 (6.67)	You can do that
Total		21 (31.34)	19 (42.22)	

Discourse Organizers

Regarding Table 10, native language users applied 40.29% of 'discourse organizer' to start talking or make rational relationships in their speech, while non-native speakers utilized 37.77% of this type. The number of 'topic elaboration/clarification bundles' in the native corpus is 16; while in non-native is 11. Both groups used sub-set "A" somehow resembling (N: 16.41% and NN: 13.33%).

Table 10

Discourse Organizers in N and NN Corpora

Functional Sets	Sub-sets	N No. (%)	NN No. (%)	Sample
2. Discourse organizers	A. topic introduction	11 (16.41)	6 (13.33)	If you think about
	B. topic elaboration/ clarification	16 (23.88)	11(24.44)	I mean it seems
Total		27 (40.29)	17 (37.77)	

Referential Expressions

As it is shown in table 11, native speakers used 'referential expressions' more than non-native speakers. It is concluded in this sub-set, 'identification/focus' was the most common. Native speakers used 10 (14.93%) 'identification/focus bundles'; non-native speakers used 3 (6.66%). Iranian learners did not employ 'imprecision bundles'; they refuse to show their doubt about the topic. On the other hand, native speakers used 1.49% of this type.

Important differences were established among the sub-sets of 'specification of attributes' native speakers utilized complicated bundles rather than Iranian English speakers. For instance, Iranian English learners used (*a lot of the*), and native speakers applied (*a little bit about*). It proved that Iranian English speakers utilized limited structures. Here in this research, number of using 'quantity of specification bundles' is equal, both two groups used 2 bundles. Neither native corpus nor non-native corpus used 'tangible framing bundles'. Regarding the 'intangible framing', just native speakers used one bundle (1.49%). 'Time reference' developed in two corpora, (NN: 2.22% and N: 2.98%). The examples are: (*at the same time, at the end of*). As the table shows, none of the learners employed 'text-deixis'. Finally, native speakers used (4.48%) 'Multi-functional reference' and non-native speakers used (4.45%) in their discussions.

Table 11

Referential Expressions in N and NN Corpora

Functional Sets	Sub-sets	N No. (%)	NN No. (%)	Sample
3. Referential	A. identification/ focus	10 (14.93)	3 (6.66)	And one of the

expressions				
	B. imprecision	1(1.49)	-	Or Sth. like that
	C. specification of attributes			
	C1. quantity of specification	2 (2.98)	2 (4.45)	A lot of things
	C2. tangible framing	-	-	In the form of
	C3. intangible framing	1(1.49)	-	Not the same thing
	D. time/ place/ text reference			
	D1. place reference	-	1(2.22)	All over the place
	D2. time reference	2 (2.98)	1 (2.22)	When you talk about
	D3. text-deixis	-	-	In the next part
	D4. multi-functional reference	3 (4.48)	2 (4.45)	At the end of
Total		19 (28.35)	9 (20)	

Discussion

The current survey scrutinized the recognized 'lexical bundles' derived from native and Iranian English Speakers' contexts. Its major benefit is proposing a bright viewpoint about the classification of linguistic aspects, as keeping formerly-recognized categorization and analogical reasons. As Zipagan and Lee's (2018) research, the outcomes of this inquiry also demonstrated that the Iranian corpus included fewer lexical bundles toward native corpus. The discoveries supported the researches (Cortes, 2004; Hyland, 2008; Ping, 2009; Shahriari Ahmadi et al., 2013; Wei & Lei, 2011) which revealed that highly developed speakers prefer to underuse bundles in their academic language comparing indigenous speakers. NNS also are inclined to apply definite bundles regularly and repeatedly, since bundles are seen "*reliable safety nets*" that can be assertively used, chiefly when they are uncertain about something (DeCock, 2000).

Conversely, some other researchers' findings contradicted the current study (Ädel & Erman, 2012; Erman, 2009). These researchers have claimed that comparing NS and NNS revealed that the second group applied a confined number of bundles. For example, Ädel and Erman (2012) investigated English highly developed authors and Swedish highly developed authors; and discovered that English authors outperformed Swedish counterparts considering both the amount and extent of bundles used.

Structural analysis, in this study, exposed that Iranian English learners utilized 'verb-phrase' and then 'noun phrase and prepositional phrase' in speaking. This result is in conformity with (Heng et al., 2014; Sykes, 2017) however Liu (2012) claimed that an overall comparison between the corpora revealed that the trend in using structural sets is very similar within both groups. Both non-native and internationally-published authors always used more phrasal bundles, most of which are noun phrase elements, across all sections of their research articles. This finding corroborates that academic writing is mainly based on noun elements (Byrd & Coxhead, 2010). Considering computer engineering, physics and applied linguistics writers; AL writers utilized lexical bundles more and the two other writers utilized 'noun phrases and prepositional phrases' more (Alipour et al., 2013).

Considering the second and third research questions in this study, findings about functional structures showed that, Iranian English speakers used 'stance expressions' more than the two other categories, while native speakers employed 'discourse organizers' more. These findings confirmed the result of the research that has been done by Kashiha and Chan (2015). They also proved that the speech of non-native speakers contained many more 'stance expressions' compared to native speakers; which was confirmed before (Bychkovska & Lee, 2017). It is also interesting to point out that, some other researchers demonstrated; 'referential bundles' and 'stance bundles' are outstanding in college orations

(Liu & Chen, 2020). On the other hand, some researchers proposed that 'topic and stance bundles' were eminent functions that were utilized in speech (Jalali & Moini, 2018; Parvizi, 2011; Valipoor, 2010).

Conclusion

The current research considered utilizing four-word bundles in speaking group discussions of native and Iranian English speakers. Throughout a corpus-based study, two corpora were contrasted according to their occurrence, category, function and also structure. The detections of the analysis manifested that the native corpus contained more lexical bundles than the Iranian corpus. The observed tendency among Iranian speakers to apply less lexical bundles has also been recorded by other researchers who have evaluated other NSs and NNSs in terms of their use of pre-fabricated patterns and formulaic language. Functionally, the findings showed that Iranian speakers used 'stance expressions' more than the two other categories, while in native corpus speakers employed 'discourse organizer' more. Structurally, in this research the evaluation exposed that learners' discussions included 'verb phrase' and then 'noun and prepositional phrase', but this is the opposite in the native corpus. This investigation expectantly is practical for teachers, students and researchers. The outcome of this study can be employed as a reference for those who desire to perform research in English teaching to build students' speaking skills and form active classes by teachers through utilizing an attractive technique for students and authentic teaching. Mostly in Iran context, the results of current study will be helpful for textbook compiler too. They can provide a list of different lexical bundles from various related studies and design creative exercises for either paragraph writing or promoting speaking skill.

Despite the remarkable results that we explored in this research, some limitations must be acknowledged. Limitations of current research contain the reasonably small number of participants and corpus size. Besides, it was restricted to speech events which occurred at two universities during a short period. The quantity of academic perspectives concentrated on is the other issue. Only one communication framework (group discussion) was observed. In order to get more generalizable findings, it sounds rational to concentrate on educational communication to achieve complete and universal outcomes.

Future studies are required to be done so as to discover the way which is helpful for learners to become familiar with bundles and employ them in their communications. Based on the current research and previous ones, other researchers can propose a table of different bundles in their teaching curriculum and speaking courses. Other researchers can investigate the way in which formulaic sequences are employed in various records, for instance, "academic lectures" and "conference presentations" in different majors and courses.

Acknowledgement

We would like to sincerely thank the editors of the Iranian Journal of English for Academic Purposes for giving us the opportunity to submit this manuscript. In addition, we would like to thank the participants for contributing to this study.

Declaration of Conflicting Interests

The authors declare that they have no conflicts of interest to disclose.

Funding Details

This research did not receive any funding.

References

Adel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, *31*(2), 81-92.

Alexander, L. G. (1970). *For and Against*. Longman group Ltd, London.

Alipour, M., Jalilifar, A., & Zarea, M. (2013). A corpus study of lexical bundles across different disciplines. *The Iranian EFL Journal*, *9*(6), 11-35.

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

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

Budiwiyanto, A., & Suhardijanto, T. (2020). Indonesian lexical bundles in research articles: Frequency, structure, and function. *Indonesian Journal of Applied Linguistics*, *10*(2), 292-303. https://doi.org/10.17509/ijal.v10i2.28592

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

Byrd, P., & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, *5*(5), 31-64.

Cancino, M., & Iturrieta, J. (2022). Assessing the impact of the lexical approach on EFL perceived oral proficiency: What is the role of formulaic sequences? *Vigo International Journal of Applied Linguistics, 19*, 41-66.

Carrol, G., & Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech*, *63*(1), 95-122.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, *23*(4), 397-423.

Crisianita, S., & Mandasari, B. (2022). The use of small-group discussion to improve students' speaking skill. *Journal of English Language Teaching and Learning*, *3*(1), 61-66.

DeCock, S. (2000). Repetitive phrasal chunkiness and advanced EFL speech and writing. In C. Mair & M. Hundt (Eds.), *Corpus linguistics and linguistic theory* (pp. 51-68). Amsterdam: Rodopi.

Erman, B. (2009). Formulaic language from a learner perspective: What the learner needs to know. In Corrigan, B., Quali, H., Moravcsik, E., & K. Wheatley (Eds.), *Formulaic language* (pp. 27-50). Amsterdam: John Benjamins.

Fajri, M. S. A., Kirana, A. W., & Putri, C. I. K. (2020). Lexical bundles of L1 and L2 English professional scholars: A contrastive corpus-driven study on applied linguistics research articles. *Journal of Language and Education*, *6*(4), 76-89. https://doi.org/10.17323/jle.2020.10719

Gablasova, D., Brezina, V., McEnery, T., & Boyd, E. (2017). Epistemic stance in spoken L2 English: The effect of task and speaker style. *Applied Linguistics, 38*(5), 613-637.

Hadizadeh, A., & Vefali, G. (2022). Formulaic language in oral academic discourse socialization of graduate students in a Northern Cyprus university. *Poznan Studies in Contemporary Linguistics*, *58*(3), 449-475. https://doi.org/10.1515/psicl-2022-0021

Heng, C. S., Kashiha, H., & Tan, H. (2014). Lexical bundles: Facilitating university "Talk" in group discussions. *English Language Teaching*, *7*(4), 1-10.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*(1), 4-21.

Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, *32*, 150-169. https://doi.org/10.1017/S0267190512000037

Jablonkai, R. (2009). 'In the light of': A corpus-based analysis of lexical bundles in two EU- related registers. *Corvinus University of Budapest: WopaLP, 3*, 1-26.

Jalali, Z. S., & Moini, M. (2018). A corpus-based study of lexical bundles discussion section of medical research articles. *Iranian Journal of Applied Language Studies*, *10*(1), 95-124.

Juknevičiene, R. (2009). Lexical bundles in learner language: Lithuania learners vs. native speakers. *Kalbotyra, 61*(3), 61-72.

Karabacak E., & Qin, J. (2013). Comparison of lexical bundles used by Turkish, Chinese, and American university students. *Procedia-Social and Behavioral Sciences, 70*, 622–628.

Kashiha, H., & Chan, S. H. (2015). A little bit about: Differences in native and non-native speakers' use of formulaic language. *Australian Journal of Linguistics*, *35*(4), 297-310.

Kwary, D. A., Ratri, D., & Artha, A. F. (2017). Lexical bundles in journal articles across academic disciplines. *Indonesian Journal of Applied Linguistics*, *7*(1), 131-140.

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. *The Journal of Asia TEFL, 11*(3), 73-103.

Lee, H. K. (2020). Lexical bundles in linguistics textbooks. *Linguistic Research, 37*(1), 121-145.

Liu, C. Y., & Chen, H. J. H. (2020). Analyzing the functions of lexical bundles in undergraduate academic lectures for pedagogical use. *English for Specific Purposes*, *58*, 122-137. https://doi.org/10.1016/j.esp.2019.12.003

Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purpose*, *31*(1), 25-35.

Maulany, D. B. (2013). The use of project-based learning in improving the students' speaking skill: A classroom action research at one of primary schools in Bandung. *Journal of English and Education, 1*(1), 30-42.

Mohammadi, M., & Enayati, B. (2018). The effects of lexical chunks teaching on EFL intermediate learners' speaking fluency. *International Journal of Instruction*, *11*(3), 179-192. https://doi.org/10.12973/iji.2018.11313

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.

Oktavianti, I. N., & Prayogi, I. (2022). Discourse functions of lexical bundles in Indonesian EFL learners' argumentative essays: A corpus study. *Studies in English Language and Education, 9*(2), 761-783.

Oktavianti, I. N., & Sarage, J. (2021). Lexical bundles in students' argumentative essays: A study of learner corpus. *Indonesian Journal of EFL and Linguistics*, *6*(2), 510-534. http://dx.doi.org/10.21462/ijefl.v6i2.421

Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes, 21*, 60-71.

Parvizi, N. (2011). *Identification of discipline-specific lexical bundles in education* [Unpublished Master's thesis]. University of Kashan, Kashan.

Pawley, A., & Syder, F.H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication,* 191-226.

Ping, P. (2009). A study on the use of four-word lexical bundles in argumentative essays by Chinese English-majors: A comparative study based on WECCL and LOCNESS. *CELEA journal*, *32*(3), 25-45.

Saadatara, A., Kiany, G., & Talebzadeh, H. (2023). Bundles to beat the band in high-stakes tests: Pedagogical applications of an exploratory investigation of lexical bundles across band scores of the IELTS writing component. *Journal of English for Academic Purposes, 61,* 101208.

Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching*. Amsterdam: John Benjamins Publishing Company. http://dx.doi.org/10.1075/scl.65

Shahriari Ahmadi, H., Ghonsoly, B., & Hossseini Fatemi, A. (2013). Analyzing research article introductions by Iranian and native English-speaking authors of Applied Linguistics. *International Journal of Research Studies in Language Learning, 2*(3), 3-18.

Shin, Y. K. (2018). *Lexical bundles in argumentative essays by native and nonnative English-speaking novice academic writers* [Unpublished Doctoral dissertation]. Georgia State University, Atlanta, GA.

Shirazizadeh, M., & Amirfazlian, R. (2021). Lexical bundles in theses, articles and textbooks of applied linguistics: Investigating intra disciplinary uniformity and variation. *Journal of English for Academic Purposes, 49*, 100946.

Simpson-Vlach, R. C., & Leicher, S. (2006). *The MICASE handbook: A resource for users of the Michigan corpus of academic spoken English*. University of Michigan Press ELT.

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*(3), 214-225.

Strunkyt, G., & Jurkūnait, E. (2008). *Written academic discourse: Lexical bundles in humanities and natural sciences* [Unpublished MA thesis]. Vilnius University, Lithuania.

Sugiarti, T. R., Fitrianasari, N. I., & Sulistyorini, T. (2018). Lexical bundles in academic writing by undergraduate and graduate students of English Language Education Program. *Loquen: English Studies Journal*, *11*(2), 1-14.

Sykes, D. L. (2017). *An investigation of spoken lexical bundles in interactive academic contexts* [Unpublished Doctoral dissertation]. Carleton University, Ottawa, Ontario.

Tavakoli, P., & Uchihara, T. (2020). To what extent are multiword sequences associated with oral fluency? *Language Learning*, *70*(2), 506-547.

Thomson, H., Coxhead, A., Boers, F., & Warren, P. (2023). Increasing use of multi-word expressions in conversation through a fluency workshop. *System*, *113*, 102994.

Valipoor, L. (2010). *A corpus-based study of words and bundles in chemistry research articles* [Unpublished Master's thesis]. University of Kashan, Kashan.

Wachidah, W. D. N. A., Fitriati, S. W., & Widhiyanto, W. (2020). Structures and functions of lexical bundles in findings and discussion sections of graduate students' thesis. *English Education Journal*, *10*(2), 131-142.

Wang, Y., & Kaatari, H. (2021). Let's say: Phraseological patterns of SAY in academic ELF communication. *Journal of English for Academic Purposes*, *54*, 101046.

Wei, Y., & Lei, L. (2011). Lexical bundles in the academic writing of advanced Chinese EFL learners. *RELC Journal*, *42*(2), 155-166.

Wingate, U. (2012). 'Argument!' helping students understand what essay writing is about. *Journal of English for Academic Purposes*, *11*(2), 145-154.

Wood, D. (2015). *Fundamentals of formulaic language: An introduction*. Bloomsbury Academic.

Wray, A., & Perkins, M. (2000). The functions of formulaic language. *Language and Communication*, *20*, 1-28.

Xu, R., & Wijitsopon, R. (2023). Corpus linguistics and cinematic discourse: Lexical bundles in mainstream film scripts. *LEARN Journal: Language Education and Acquisition Research Network*, *16*(1), 545-574.

Xuan, C., & Kim, Y. J. (2022). A comparative study of lexical bundles between English education and linguistics. *The Journal of Modern British & American Language & Literature*, *40*(2), 233-254.

Yakut, I., Yuvayapan, F., & Bada, E. (2021). Lexical bundles in L1 and L2 English doctoral dissertations. *Journal of Teaching English for Specific and Academic Purposes*, *9*(3), 475-493.

Yang, B. (2020). A corpus-based study of synonymous epistemic adverbs perhaps, probably, maybe and possibly. *Research Journal of Education*, *6*(8), 158-168.

Yang, W., & Fang, Q. (2021). The structures and functions of lexical bundles in argumentative essays by Chinese EFL students at the tertiary level. *International Journal of TESOL Studies*, *3*(3), 41-60.

Yang, Y. (2017). Lexical bundles in argumentative and narrative writings by Chinese EFL learners. *International Journal of English Linguistics*, *7*(3), 58-69.

Zipagan, M. N., & Lee, K. R. (2018). Korean English learners' use of lexical bundles in speaking. *Journal of Asia TEFL*, *15*(2), 276-291.

Chabahar Maritime University
Iranian Journal of English for Academic Purposes ISSN: 2476-3187
IJEAP, 2023, 12(3), 56-72 (Previously Published under the Title: Maritime English Journal)

Appendix

List of Extracted Lexical Bundles from Non-native Speakers' Corpus

At the end of	Can be used as	A little bit more
For a long time	Do you know that	Do you need to
I think it is	For the sake of	I think the best
I mean it seems	I'm going to	If you want to
I guess you can	To come up with	When we want to
One of the most	What do you mean	It seems to everyone
I was just wondering	So I want to	That would be good
Let's talk about the	To be able to	When it comes to
This is hard to	You talk about the	Where do you prefer
I think we have to	I agree with you	If you think about
I have no idea	In the same way	If we want to
This is a little	It is hard to	That would be interesting
It's your turn	On the other hand	Thank you very much
I wanted to talk	It seems to be	Some years ago when
What is your idea	What that means is	What do you think