



## A Corpus based Study of Lexical Bundles in Native and Non-Native Speaking Group Discussions

Ashraf Vaziri<sup>1</sup>, Hamed Barjasteh<sup>2\*</sup>, Atefeh Nasrullahi Mozirji<sup>3</sup>

<sup>1</sup>Ph.D. Candidate, Department of English Language and Literature, Islamic Azad University, Ayatollah Amoli Branch, Amol, Iran

<sup>2\*</sup>Assistant Professor, Department of English Language and Literature, Islamic Azad University, Ayatollah Amoli Branch, Amol, Iran

<sup>3</sup>Assistant Professor, Department of English Language and Literature, Islamic Azad University, Ayatollah Amoli Branch, Amol, Iran

Received: August 05, 2023

Accepted: October 23, 2023

### Abstract

Speech is unquestionably organized with lexical bundles that are managed along with syntactic rule and hold each own mixture of functions. Lexical bundles make written and spoken language more coherent, it has also crucial position in the comprehension of context. So, investigating lexical bundles in different circumstances is beneficial for learners in creating suitable speech. This research considered the “frequency”, “structure” and “function” of four-word lexical bundles in speaking group discussions. British National Corpus (BNC) was applied as a native corpus in this research and Iranian intermediate EFL learners were the non-native corpus. To achieve this goal, by using AntConc 3.3.0 software, four-word lexical bundles were extracted from twenty-one group discussions corpus and were classified by employing Biber et al. (2004) “structural and functional” taxonomies. Comparing both corpora, outcomes showed that *Native Speakers* utilized more “lexical bundles” than *Non-native*. Besides *Native Speakers* employed more “discourse organizing bundles” in functional classification and “verb phrases” in structural classification, whereas *Non-native Speakers* more normally utilized lexical bundles as “stance expressions” in functional classification and “verb phrase fragments” in structural classification. The results of the research have some significant educational indications for language EFL teachers, researchers and learners too.

**Keywords:** Corpus analysis, Lexical bundles, Speaking group discussion

### INTRODUCTION

Of all four main language skills, speaking is taken into account to be the foremost consequential in learning a second or foreign language. For Brown and Yule (1983), speaking is the skill that the learners are going to be assessed most in real-life contexts. Meanwhile, speaking is the ability of talking and conveying thoughts by using language (Mart, 2012). Sihotang et al. (2021) proposed that “Speaking

is a basic skill that should be possessed to communicate” (p. 5). A lot of researches have been done to assist learners’ speaking. Still many EFL learners believe that speaking is difficult (Cancino & Iturrieta, 2022; Zipagan & Lee, 2018). Besides, Hinkel (2005) explained speaking as “the most complex and difficult skill to master” (p. 485). Mohammadi and Enayati (2018) declared that learning speaking skill assertively seems as a challenge for most English students, even if students waste lots of time and funds for gaining this skill. In order to

\*Corresponding Author’s Email:

[h.barjasteh@iauamol.ac.ir](mailto:h.barjasteh@iauamol.ac.ir)

boost speaking skill, Crisianita and Mandasari (2022) stated that accuracy based activities such as: *discussion, speeches, role-plays and conversations* can be helpful. In support with their idea, Fauzi (2017) proposed that through speaking and listening in group discussions and receiving effective correctness, learners can improve their speaking skill. Furthermore, topical knowledge, vocabulary knowledge, self-confidence and other factors can impact on learners' speaking skill (Bohari, 2020). According to Unsworth and Mills (2020) English speakers may hardly elucidate without syntax, but nothing can be expressed without lexis.

The explorers of the lexical perspective, Mumford and Dikilitaş (2020) claimed that the construction of language connection are not syntax, semantics, concept, or further elements like teaching strategies; except vocabulary and word configurations. In other words, multi-word constituents have essential role in spoken language, so many researchers started to investigate the expressions in constructing fluency and confidence (Verzella, 2020). Generally, among "lexical bundles"; two-word series are utilized too many and five- or six-word sequences are inversely (Biber, 2014). One of the most important features in describing "lexical bundles" is their "range" – that the quantity of contents within that they are utilized in corpus. So we cannot say every series of four-words is applicable and authentic. Just continuous strings of vocabularies are regarded "lexical bundles"; any component that passes restrictions on use or a punctuation mark would be kept out (Biber et al., 2004; Budiwiyanto & Suhardijanto, 2020). Not only "lexical bundles" make the structures of sentences perfect, but they also have a fundamental role in speech, they act like: "a kind of pragmatic head for larger phrases and clauses, where they function as discourse frames for the expression of new information" (Biber & Barbieri, 2007, p. 270). In addition to conversation patterns EFL learners use multi-word chunks to improve their speaking and being native-like speakers (Khoiriyah & Mujiyanto, 2022). Furthermore, some researchers like (Conklin & Schmitt, 2008; Wood, 2006; Wray & Fitzpatrick, 2008) believed that lexical bundles are conventionalized in a definite

register, so comprehending features of them will be valuable for L2 speakers as a means to boost their speaking skills. Besides, Tang (2012) came to the conclusion that representing lexical bundles for learners can enhance their accuracy and authenticity in their talking (As cited in Cancino & Iturrieta, 2022). Carpenter et al. (2020) and some other researchers claimed, research on lexical bundles has the supposition that fluent Native Speakers employ plenty of bundles in daily life language (Conklin & Schmitt, 2008; Indriyani et al., 2022).

As Granger (1998) stated, since there aren't enough appropriate instruction for utilizing the lexical bundles, EFL learners have problem with using the bundles, and they just have the chance in classroom context. One way to accomplish this problem is using the researches. As Albelihi (2022) stated "lexical chunks" are rarely conferred in systematized linguistics. But, there are some studies which centralized on written context in order to investigate various features of "lexical bundles" (Abdaloussein, 2022; Ädel & Erman, 2012; Hyland, 2008; Shirazizadeh & Amirfazlian, 2021; Vo, 2019). And there are few researchers related to utilizing lexical bundles in speaking context (Cancino & Iturrieta, 2022; Crisianita & Mandasari, 2022; Nesi & Basturkmen, 2009). While, discerning the knowledge of lexical bundles that are employed in speaking as the framework for undergraduates in acquiring English as a foreign language is very significant. In spite of the significance and recurrent applying of "lexical bundles", there are still main issues related to their specific utilization in group discussions. In context of Iran, "lexical bundles" have not been extensively disserted, thus to fill this breach, current study aimed at investigating "lexical bundles" in corpus of "*group discussion sessions*". Elicited bundles were considered both structurally and functionally. After that, Iranian corpus was compared with native corpus. The researchers chose the corpus-based study for the reason that corpus technology brings in new ways of accessing information and presentation, facilitating both the processes of encoding and decoding. On the other hand, the bundles which recognized in

this research may be beneficial for inexperienced teachers or learners and also EAP teachers to prioritize these formulaic expressions in syllabus designing and lesson plan.

In order to solve this problem, the subsequent research questions were posed in this research:

**RQ1.** Which lexical bundles are utilized generally in speaking of English language Institute learners' group discussions?

**RQ2.** What are the "structural and functional" characteristics of extracted lexical bundles in the institute group discussions among English language learners?

**RQ3.** What are the resemblances and distinctions between two corpora, considering bundle' "occurrence", "structure" and "function"?

## LITERATURE REVIEW

### Lexical Bundles

An essential portion of language learning is the aptitude to utilize formulaic sequences, preassembled series of vocabularies which contain definite structure, connotation and operation. Considering multi-word combinations has attracted the attention of researchers for a long time. Researchers utilized several terms mentioning multi-word combinations. Jeperesen (1924) and Firth (1957) were one of the first linguists that noticed "multi-word expression" and called them "collocation" (As cited in Karima Ibrahim, 2019). Throughout years, word series have been dominated differently: "prefabs" (Bolinger, 1976, as cited in Shirazizadeh & Amirfazlian, 2021), "lexical phrases" (Nattinger & DeCarrico, 1992, as cited in Shirazizadeh & Amirfazlian, 2021), "lexical bundles" (Biber et al., 2004). One description regarding "lexical bundle" is:

"A sequence, continuous or discontinuous, of words or other meaning elements, which are or appear to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar" (Wray, 2000, p. 465).

Through investigating of "lexical bundles", some researchers did essential work. Actually

"Lexical bundles" are not established linguistic components; they have a practical source depended on the occurrence criterion (Lee, 2020; Salazar, 2014). The result of the investigation of Biber et al. (2004) indicated, most "formulaic sequences" link two clauses which have consistent grammatical connection and can be recognized as "Structural expressions". They proposed three chief *structural* forms: "verb phrases fragments", "clause fragments", and "noun and preposition phrase fragments". Biber et al. also proposed *functional* taxonomy which has three categories: "stance bundle", "discourse organizer", and "referential expression". To be more analytical,

"*Stance bundles* express attitudes or assessments of certainty that frame some other proposition; *discourse organizers* reflect relationships between prior and coming discourse; *referential bundles* make a direct reference to physical or abstract entities, or to the textual context itself, either to identify the entity or to single out some particular attribute of the entity as especially important" (Biber et al., 2004, p. 384).

Another perception of "lexical bundle functions" was recommended by Hyland (2008). He recommended special categorization that concentrated on the characteristics of written context and categorizes the bundles into "research-oriented bundles", "text-oriented bundles", and "participant-oriented bundles". There are some researches that used Hyland taxonomy for analyzing learners' essays or written articles (Al et al., 2020; Liu & Chen, 2020; Yakut et al., 2021). But in this research, the researcher employed the first taxonomy.

### Previous Studies

Several researches considered the discords in "lexical bundle" usage among writers (Navarro Gil & Martínez Caro, 2019; Pan et al., 2016; Salazar, 2014) and some researches concentrated on "theses or dissertations" (Wachidah et al., 2020; Yakut et al., 2021) "published articles" (Budiwiyanto & Suhardijanto, 2020) or "books" (Hussain et al., 2021). These investigations are significant to explain the intellectual learners' language competence or to recognize

peculiar expressions in particular subjects. Besides, the lack of appropriate instruction concerning the use of formulaic strings in EFL classrooms is frequently observed and consequently, EFL learners who are generally far from native speaker environments and their language learning opportunities are limited to class hours undergo the deficiency of instruction (Granger, 1998).

As mentioned before, there are few studies related to speaking and lexical bundles. To name some we can talk about the subsequent researches. Investigating '160 academic lectures', researchers have inferred that "lexical bundles" have a momentous impress in developing cohesion (Nesi & Basturkmen, 2009). Shokri (2010) found how lexical bundles had helped learners to develop their communication skill, strategies, and self-confidence. Shokri's survey recognized that learners have positive approach toward a project. Research has been carried on the examination of lexis familiarity in short talks. It can determine the word models, aspects, and application that employed by learners in real life sociability and particular corpus TED was organized (Wang, 2012). Examining the impact of "collocations" on Iranian students' speaking was done by Attar and Allami (2013) and they realized that training lexical collocation was practical and improve collocation knowledge and speaking skill. According to the results of Mahdavi-Zafarhandi and Emamzadeh (2016) research, utilizing readymade bundles in the situation of classroom increase students' fluency in speaking. When they became attentive of the benefits of employing "lexical bundles", they persuaded to apply them. McCarthy and Carter (2004) also proved that employing bundles can boost fluency and coherence. Zipagan and Lee (2018) found that Korean English learners prefer to employ "lexical bundles" in their speech more willingly rather than writing. The other study about "academic lectures" revealed that both "referential bundles" and "stance bundles" are outstanding within them (Liu & Chen, 2020).

## METHOD

### **Iranian Corpus**

Iranian Non-native spoken corpus included 21 group discussion transcripts containing

various subjects like: "how to control the stress, how to promote English skills, Traveling". The contributors of the group discussions were 28 female and male intermediate EFL learners (divided in 7 group discussions) with age range of 14-18 studying in Danesh Institute in Tehran. All participants selected for the study got at least 90 score out of 100 in their last three final exams. The participants were selected based on 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**

The "British National Corpus" (BNC) was chosen for this study.

"Except written texts, it contains 10 million spoken words. It is a finite, balanced, sampled corpus made up two kinds of spoken English, conversational – males and females, aged 15–60+, from different social groups and regions, and task-oriented – lectures, company and trade union talks, business meetings, consultations, sermons, parliamentary and legal proceedings, TV/radio broadcasting, speeches, commentaries, etc" (Grant, 2005, p. 438).

### **Corpus Taxonomy**

The current study utilized "functional and structural" categories of Biber et al. (2004), since their research has been well considered analyzing spoken data. In line with "structural classification", there are three chief grammatical forms: "1) lexical bundles which contain phrase portions; 2) lexical bundles which contain subordinate clause fragments; and 3) lexical bundles which include phrase and prepositional phrases" (p. 381). They also offered three main "functional classification":

"Stance bundles are described because the overt expression of an author's or speaker's attitudes, feelings, judgments, or commitment concerning the message; discourse organizer bundles demonstrate the correlations between given and coming discourse; and referential bundles create straight relevance

of physical or abstract entities” (p. 384).

### Corpus Tool

The analysis in this study was achieved using a software program AntConc 3.3.0 developed by Anthony (2019). The software presents its users with the following analysis tools: “concordance, concordance plot, files view, clusters, N-grams, collocates, word list and keyword list”. The “N-gram” of the software was employed to distinguish existent “lexical bundles” in every corpus throughout an experimental analysis.

### Procedure

The participants were assigned to have discussion based on the specific topics and their discussions have been recorded by the researcher. Then their discussions have been transcribed to *Notepad* format. In the next phase, the researcher applied AntConc 3.3.0, freeware concordance computer software to categorize and establish a list of common four-

word expressions in sub-corpus. In this stage “occurrence cut-off point” for the detection of bundles should have been set, for this study it would be four occurrences in corpus. After extraction the lexical bundles, they were classified in to “structural and functional” categories utilizing the taxonomies of Biber et al. (2004) and the researcher compared two corpora.

## RESULTS

### Distribution of the Target Bundles

First analysis of two corpora (native speakers (NSs) and non-native speakers (NNSs)) of extracted bundles exposed that native speakers' discussions comprised more bundles compared to NNS. On the whole, total of 193 target bundles were found in native corpus and 151 bundles in Iranian English Speakers corpus. The results revealed that 43 varied bundles were in native corpus and 32 in NNSs corpus. In fact that second language English speakers made use of fewer “lexical bundles” than native speakers has been proved by (Kashiha & Chan, 2015) in the previous studies.

**Table 1**

*Division of Lexical Bundles in N and NN Corpora*

Groups	Entire No. of N-grams types	Entire No. of N-grams Token
Native speakers	43	193
Non- native Speakers	32	151

*Note.* N-gram: four-word lexical bundles

### Structural Division of Lexical Bundles

Analyzing results explained bundles mostly included “verb phrases” in both N and NN corpora. Table 2 provides the number and percent of the most important syntactic forms of the bundles within discussions. As it is specific in the table below, native speakers were used

“verb phrase” more than the two other categories. NSs used “verb phrase” (55.82%), “noun and prepositional phrase” (30.23%) and “dependent clause” (13.95%). While; non-native speakers utilized “verb phrase” (50%) and “noun and prepositional phrase” (34.38%) and “dependent clause” (15.62%).

**Table 2**

*Structural Division of Lexical Bundles in N and NN Corpora*

Structural types	N No. (%)	NN No. (%)
Verb phrase	24 (55.82)	16 (50)
Noun phrase and Prepositional phrase	13 (30.23)	11 (34.38)
Dependent clause	6 (13.95)	5 (15.62)
Total	43	32



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

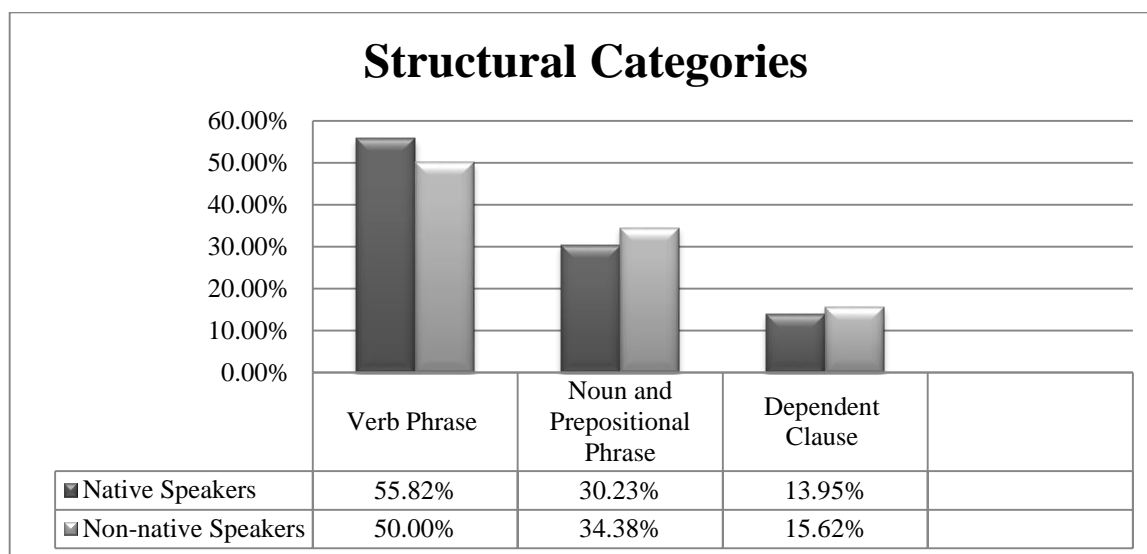


Figure 1

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

### Verb Phrase

As it is explained in this part, the number of “verb phrase” in two corpora is more than the other two sub-categories in structural classification. In non-native corpus, speakers used “first/ second person pronoun+ VP” (9.37%), “discourse marker + VP” (9.37%), “yes/no question” (9.37%) and “Wh-question fragment” (9.37%) more than other sub-categories.

Whereas, in native corpus speakers employed “third person pronoun + VP” (11.62%), “first/ second person pronoun+ VP” (9.31%), “yes/no question” (9.31%) and “Wh-question fragment” (9.31%) more than other sub-categories. Based on the findings, it’s interesting to explain that both corpora are somehow employed the same sub-set bundles in verb phrase.

Table 3

*VP in N and NN Corpora*

Structural Sets	Sub-sets	N No. (%)	NN No. (%)	Sample
Lexical bundles that incorporate VP	a. 1 <sup>st</sup> /2 <sup>nd</sup> person pronoun + VP	4(9.31)	3 (9.37)	I don’t want to
	b. 3 <sup>rd</sup> person pronoun + VP	5 (11.62)	2 (6.25)	This is one of
	c. discourse marker + VP	3 (6.97)	3 (9.37)	I think this is
	d. VP (with non-passive verb)	2 (4.65)	1 (3.13)	Let’s look at the
	e. VP (with passive verb)	2 (4.65)	1(3.13)	Has been proved that
	f. yes/ no question	4 (9.31)	3 (9.37)	Do you agree with
	g. Wh-question	4 (9.31)	3 (9.37)	What do you mean
Total		24 (55.82)	16 (50)	

### Dependent Clause

As table 5 demonstrates, using dependent clause was the third and better to say the last sub-category of structural lexical bundles that both NSs and NNSs have made use of in their discussions. NSs used 6 (13.95%) out of 67 four- word lexical bundles. Both NSs and NNSs used “first/second person pronoun + dependent clause” as the most common sub-category (for

example: I don’t know about). For native corpus it is 4.65% and for non-native corpus 6.25%. The second sub-category for native corpus is “Wh-clause” which is again 4.65%. Using bundles in other sub-categories except “that clause” in non-native corpus are the same (3.125%). It is essential to declare that neither NSs nor NNSs employed “that clause” in their discussions.

**Table 5**  
**Dependent Clause in N and NN Corpora**

Structural Sets	Sub-sets	N No. (%)	NN No. (%)	Sample
Lexical bundles that incorporate dependent clause	a. 1 <sup>st</sup> /2 <sup>nd</sup> person pronoun +dependent clause	2 (4.65)	2 (6.25)	I don't know whether
	b. wh-clause	2 (4.65)	1(3.125)	When you say that
	c. if-clause	1 (2.32)	1(3.125)	If you think about
	d. (verb/adjective) + to-clause	1 (2.32)	1(3.125)	To think how to
	e. that clause	-	-	Is that there is
Total		6 (13.95)	5 (15.62)	

As the results revealed, the performance of NS and NNS within using “verb phrase” of “structural bundles” are the same. The greater part of “lexical bundles” in the corpus is, “phrasal” more willingly than “clausal” in proportion to findings of prior researches (Bal, 2010).

#### Functional Division of Lexical Bundles

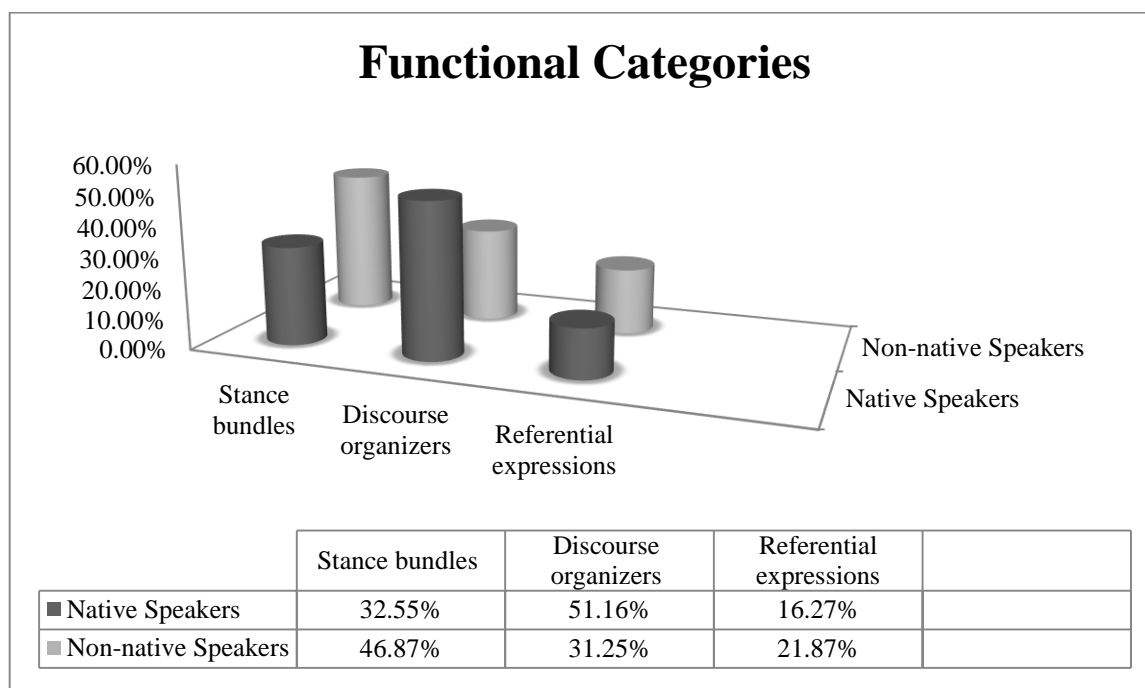
The results prove that spoken language in group discussions in N corpus is expressed by 22 “discourse organizers” (51.16%), go after by 14

“stance expressions” (32.55%), 7 “referential expressions” (16.27%). Juknevičienė (2009) claimed that “stance expressions” and “discourse organizing” could be used in spoken context more, while “referential expression” is employed in written form. It can be realized that Iranian English speakers employed 46.87% of “stance expressions”. In contrast, native speakers revealed more propensities to use 51.16% of “discourse organizers”, that was 31.25% in Iranian corpus. Accidentally, two corpora had related proportions in the employment of “referential expressions”.

**Table 6**  
**Functional Division of Lexical Bundles in N and NN Corpora**

Functions	N No. (%)	NN No. (%)
Stance expressions	14 (32.55)	15 (46.87)
Discourse organizers	22 (51.16)	10 (31.25)
Referential expressions	7 (16.27)	7 (21.87)
Total	43	32

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**

As it is illustrated in Table 7, functionally most of the “lexical bundles” employed by Iranian students were “stance bundles”, they used 15 out of 32; with the percentage of (46.87%). It is clear that L2 speakers’ tendency is using “stance bundles” as an important function. This result is quite the opposite of native speakers’ inclination. However, in “stance bundles”, both corpora employed “epistemic stance expressions” more than other sub-categories, (11.62% and 15.62% respectively).

One phrase that is employed by EFL learners more is “I don’t think that” which confirmed 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”. “Intention/Prediction” subset was more common in both native (6.98%) and non-native (9.375%) corpus. But there is discrepancy in the structure of expressions utilized. Non-native speakers utilized phrases like: “I am going to”, to intend the plan. Native speakers employed complicated utterances, like: “I was expected to”.

For “ability expressions” NSs were proficient to employ “will be able to + verb” structure which is inferred “double hedge” and specifies capability more openly. In “obligation/directive” sub-set, Iranian speakers appeared to apply these series to direct one another in discussing, or to give prominence to the significance of an occasion, like: “you have to do”. NSs use this function, but it is not very general their talking.

**Table 7**

*Stance Bundles in N and NN Corpora*

Functional Sets	Sub-sets	N No. (%)	NN No. (%)	Sample
Stance bundles	A. epistemic stance	5 (11.62)	5 (15.62)	I think it was
	B. attitudinal/modality stance			
	B1. desire	1 (2.32)	1 (3.125)	I don’t want to
	B2. obligation/ directive	2 (4.65)	3 (9.375)	Do you have to
	B3. intention/prediction	3 (6.98)	3 (9.375)	I am going to
	B4. ability	3 (6.98)	3 (9.375)	You can do that
Total		14 (32.55)	15 (46.87)	



### Discourse Organizers

Concerning Table 8, “discourse organizer bundles” were employed in two corpora, either to commence talking, or create normal connection linking opinions. Native speakers used 51.16% “discourse organizers” in their discussions, though non-native speakers utilized 31.25% of this sort. Truly as the researcher stated in the preceding section, the most repeated functional bundles that L2 speakers applied, was “stance bundles”. The quantity of “topic elaboration/clarification bundles” in native corpus is 14 (32.55%), as in non-native are 7 (21.875%).

“Topic introduction/focus bundles” were used by NSs (18.61%) and Iranian speakers (9.375%). Actually, the speakers in both corpora

utilized these bundles in order to begin talking in their discussions, or concentrate on the subject. Furthermore, there were distinctions between the kinds of series and the technique they were employed. For instance, NSs apply formal bundles in compare with non-native speakers. “I would like to” is one of those formal bundles. Possibly the cause of using formal structures by L1 speakers is because of their type of education. By taking into consideration the result, the other varied feature in natives and non-natives talking construction is the employment of the “question bundle”: “what is your opinion?” It was utilized to ‘*start a topic*’ by native speakers; and ‘*at the end of their talks*’ by Iranian students.

**Table 8**

*Discourse Organizers in N and NN Corpora*

Functional Sets	Sub-sets	N No. (%)	NN No. (%)	Sample
Discourse organizers	A. topic introduction	8 (18.61)	3 (9.375)	If you think about
	B. topic elaboration/ clarification	14 (32.55)	7 (21.875)	I mean it seems
Total		22 (51.16)	10 (31.25)	

### Referential Expressions

As it is shown in table 9, the first “referential expressions” that the speakers used were “identification/focus bundles”. Native speakers used “referential expressions” more than L2 speakers. In fact, speakers employ these kinds of bundles with the aim of referring to accentuate something. Native speakers applied (4.65%) “identification/focus bundles”; non-native speakers used (3.125%). Applying “imprecision bundles” did not come out within both corpora that would typically be presumed in offhand talking. Mostly, native speakers employed the bundle “or something like that” to point to a preceding reference in academic context. In this study neither native nor non-native speakers utilized this kind of bundle in general context. Generally, most non-native speakers may decline use this kind of bundle, since they try to avoid expressions that show their uncertainty about the topic of discussion.

There are significant distinctions among the sub-sets of “specification of attributes”; native speakers apparently utilized bundles in different quantities, rather than simpler expressions like: “a lot of the”, which were used by Iranian English speakers. The model that natives used is “a little bit about”. It confirmed that Iranian English speakers utilized restricted bundles. In this research the number of using “quantity of specification bundles” are somehow equal, L1 speakers used (4.65%) and L2 speakers used (9.375%). Neither native corpus nor non-native corpus employed “tangible framing and intangible framing” bundles in general context. “Time reference bundles” emerged in native corpus (4.65%) and non-native speakers (3.125%), like: “at the same time”, “at the end of”. Based on the findings none of the speakers used the “text-deixis bundles”. Noticeably, this function is not applicable in spoken context. Lastly, native speakers used (2.32%) “multi-functional reference” and non-native speakers used (6.25%) in their discussions.

**Table 9**  
**Referential Expressions in N and NN Corpora**

Functional Sets	Sub-sets	N No. (%)	NN No. (%)	Sample
Referential expressions	A. identification/focus	2 (4.65)	1 (3.125)	And one of the
	B. imprecision	-	-	Or sth. like that
	C. specification of attributes			
	C1. quantity of specification	2 (4.65)	3 (9.375)	A lot of things
	C2. tangible framing	-	-	In the form of
	C3. intangible framing	-	-	Not the same thing
	D. time/place/text reference			
	D1. place reference	-	-	All over the place
	D2. time reference	2 (4.65)	1 (3.125)	When you talk about
	D3. text-deixis	-	-	In the next part
	D4. multi-functional reference	1 (2.32)	2 (6.25)	The end of the
	Total	7 (16.27)	7 (21.87)	

## DISCUSSION

In this study, researcher considered “frequency”, “structure”, and “function” of four-word LBs of Ns and NNs general speakers within group discussions. Related to the first research question which considered “frequency”, native speakers applied a higher number of bundles than Iranian English Speakers. The current finding is in line with some preceding researches like (Kashiha & Chan, 2015; Zipagan & Lee, 2018), outcomes. In the first study, Kashiha and Chan contrasted four-word lexical bundles in “*spoken educational discourse of native and Malaysian English learners*”. The researchers in the second example considered the “frequency; structures and functions” of lexical bundles utilized by Korean English students and found that in comparing advanced and novice learners, advanced learners are less dependent on using lexical bundles in their speaking (Zipagan & Lee, 2018). Yoon and Choi (2015) also supported Zipagan and Lee’s result. In opposition, it’s interesting to mention that Jones et al. (2015) analyzed “lexical bundles” in spoken language and explored that utilizing “lexical bundles” is more usual at upper proficiency levels.

There are also some researches which are in contrast with the current result in this study (Ädel & Erman, 2012; Shahmoradi et al., 2021; Shahriari Ahmadi et al., 2013; Wei & Lei, 2011). For instance, Shahriari Ahmadi et al. (2013) evaluated “Applied Linguistics” research

papers in both native and non-native corpus; their detections exposed that Iranian writers utilized more 4-word lexical bundles. Shahmoradi et al. (2021) corroborated Shahriari Ahmadi et al. discovery, since they investigated English articles in Applied Linguistics and Information Technology and considered native authors with non-native authors. Their conclusions affirmed that native authors applied less lexical bundles comparing Iranian authors.

Referring to second and third research question structurally, both native and non-native speakers employed “verb phrase” and “noun phrase and prepositional phrase” rather than dependent clause bundles. Using “verb phrase” in this study is in line with (Heng et al., 2014; Sykes, 2017; Yousaf & Shehzad, 2018). Considering “lexical bundles in group discussions” performed by skillful ESL university students showed that learners employed “verb phrase” more than the other two sub-categories (Heng et al., 2014). Investigating PhD dissertations composed by Pakistani learners proved that “verb phrase” was the prominent category in the corpus of Bio Sciences (Yousaf & Shehzad, 2018). On the other hand, Cui and Kim (2023) stated that “noun phrase-based” and “preposition phrase-based” bundles were the eminent ones and it is confirmed by (Hyland & Jiang, 2018; Shirazizadeh & Amirfazlian, 2021). Oktavianti and Sarage (2021) declared English learners at university level employed more “noun phrase” structures in their writing essays.

Assessing functionally to answer to second and third research question, the detections in the Iranian corpus proved, the majority of bundles in speaking context are consisted of: “stance bundles, discourse organizers, referential bundles” respectively. This result supported (Jalali & Moini, 2018; Valipoor, 2010; Zipagan & Lee, 2018) discoveries. Xu and Wijitsopon (2023) examined the wording of American mainstream film scripts, and found that they are characterized generally by spoken formulaic expression and descriptive expressions, like “place-referential” and “action-related” lexical bundles. Karima Ibrahim (2019) in observing academic lectures proved that “referential bundles” were the most general bundles in non-native corpus.

Although the researcher achieved important outcomes in the current study, there were some limitations. The first one can be the size of the corpus, as it was mentioned before this research considered only twenty-one group discussions. The second one is the matter of investigating just speaking group discussions of the non-native corpus, other speech contexts can be considered in the further researchers.

## CONCLUSION

To summarize, Iranian English learners group discussion were considered and 4-word lexical bundles extracted and categorized. After that the findings of the research were compared with the BNC native corpus. Regarding the frequency, non-native learners used bundles less than native counterparts. In structural and functional groups, they employed “verb phrase” and “stance bundles” respectively. Considering pedagogically, the results of this investigation are practical for material designers or teachers for speaking and should comprise them more significantly in their materials and syllabus. Providing English learners with the “lexical tools” that will bring them closer to the native norm appears to be an intellectual purpose. 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 lexical bundles are employed in various records, for instance, “academic lectures” and “conference presentations” in different majors and courses in context of Iran or other countries.

## References

- Abdaloussein, H. F. (2022). Iraqi EFL learners' use of formulaic language in writing proficiency exams. *Journal of Language and Linguistic Studies*, 18(1), 1079-1093.
- Ädel, 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.
- Al, F. M. S., Wasito, K. A., & Kharisma, P. C. I. (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 (24)), 76-89.
- Albelihi, H. H. M. (2022). Intermediate English as a Foreign Language learners' formulaic language speaking proficiency: Where does the teaching of lexical chunks figure? *Frontiers in Psychology*, 13, 949675.
- Anthony, L. (2019). *AntConc (Version 3.5.8)[Computer Software]*. Waseda University. In
- Attar, E. M., & Allami, H. (2013). The effects of teaching lexical collocations on speaking ability of Iranian EFL learners. *Theory and Practice in Language Studies*, 3(6), 1070-1079.
- Bal, B. (2010). Analysis of Four-word Lexical Bundles in Published Research Articles Written by Turkish Scholars.
- Biber, D. (2014). Intra-textual variation within medical research articles. In *Variation in English* (pp. 108-123). Routledge.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for specific purposes*, 26(3), 263-286.

- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied linguistics*, 25(3), 371-405.
- Bohari, L. (2020). Improving speaking skills through small group discussion at eleventh grade students of SMA Plus Munirul Arifin NW Praya. *Journal of Languages and Language Teaching*, 7(1), 68-81.
- Brown, G., & Yule, G. (1983). *Teaching the spoken language* (Vol. 2). Cambridge university press.
- Budiwiyanto, A., & Suhardijanto, T. (2020). Indonesian lexical bundles in research articles: Frequency, structure, and function. *Indonesian Journal of Applied Linguistics*, 10(2).
- Cancino, M., & Iturrieta, J. (2022). Assessing the impact of the Lexical Approach on EFL perceived oral proficiency: What is the role of formulaic sequences?
- Carpenter, S. K., Witherby, A. E., & Tauber, S. K. (2020). On students' (mis) judgments of learning and teaching effectiveness. *Journal of Applied research in Memory and cognition*, 9(2), 137-151.
- Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied linguistics*, 29(1), 72-89.
- 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.
- Cui, X., & Kim, Y. (2023). Structural and functional differences between bundles of different lengths: A corpus-driven study. *Frontiers in Psychology*, 13, 1061097.
- Fauzi, I. (2017). Improving students' speaking ability through small-group discussion. *Journal of ELT Research*, 2(2), 130-138.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and lexical phrases. *Phraseology: Theory, analysis and applications*, 145-160.
- Grant, L. E. (2005). Frequency of 'core idioms' in the British National Corpus (BNC). *International Journal of Corpus Linguistics*, 10(4), 429-451.
- Heng, C. S., Kashiha, H., & Tan, H. (2014). Lexical Bundles: Facilitating University "Talk" in Group Discussions. *English Language Teaching*, 7(4), 1-10.
- Hinkel, E. (2005). *Handbook of Research in Second Language Teaching and Learning*. New Jersey: Lawrence Elbaum Associates. Inc.
- Hussain, G., Zahra, T., & Abbas, A. (2021). Discourse Functions of Lexical Bundles in Pakistani Chemistry and Physics Textbooks. *GEMA Online Journal of Language Studies*, 21(1).
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for specific purposes*, 27(1), 4-21.
- Hyland, K., & Jiang, F. (2018). Academic lexical bundles: How are they changing? *International Journal of Corpus Linguistics*, 23(4), 383-407.
- Indriyani, L., Rukmini, D., & Widhiyanto, W. (2022). The realization of formulaic expressions in EFL teacher-student classroom interactions. *English Education Journal*, 12(1), 121-130.
- Jalali, Z. S., & Moini, M. (2018). A corpus-based study of lexical bundles in discussion section of medical research articles. *Iranian Journal of Applied Language Studies*, 10(1), 95-124.
- Jones, C., Waller, D., & Golebiewska, P. (2015). Spoken lexical chunks used by successful learners at B2 level: Forms and functions. *CALR Linguistics Journal*, 1-13.
- Juknevičienė, R. (2009). Lexical bundles in learner language: Lithuanian learners vs. native speakers. *Kalbotyra*, 61, 61-72.
- Karima Ibrahim, E. T. (2019). *A study of lexical bundles in academic lectures in the base corpus/Karima Ibrahim El Twati University of Malaya*.
- 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.
- Khoiriyah, L., & Mujiyanto, J. (2022). The realization of formulaic competence in the classroom interactions among learners in Kampung Inggris Pare. *English Education Journal*, 12(2), 141-150.
- Lee, H.-K. (2020). Lexical bundles in linguistics textbooks. *Linguistic Research*, 37(1).
- 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.
- Mahdavi-Zafarghandi, A., & Emamzadeh, S. (2016). The effect of teaching lexical collocations on Iranian EFL learners' writing ability: focusing on the appropriate use of collocations. *Journal of Applied Linguistics and Language Research*, 3(5), 107-117.
- Mart, C. T. (2012). Developing speaking skills through reading. *International Journal of English Linguistics*, 2(6), 91.
- McCarthy, M., & Carter, R. (2004). This that and the other: Multi-word clusters in spoken English as visible patterns of interaction. *TEANGA, the Journal of the Irish Association for Applied Linguistics*, 21, 30-52.
- 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.
- Mumford, S., & Dikilitaş, K. (2020). Pre-service language teachers reflection development through online interaction in a hybrid learning course. *Computers & Education*, 144, 103706.
- Navarro Gil, N., & Martínez Caro, E. (2019). Lexical bundles in learner and expert academic writing. *Bellaterra journal of teaching and learning language and literature*, 12(1), 0065-0090.
- Nesi, H., & Basturkmen, H. (2009). Lexical bundles 121 and discourse signaling in academic lectures. *Flowerdew J. Malhberg M. (Eds.), Lexical Cohesion and Corpus Linguistics*, 23-44.
- 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.
- 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.
- Salazar, D. (2014). Lexical bundles in native and non-native scientific writing. *Lexical Bundles in Native and Non-native Scientific Writing*, 1-221.
- Shahmoradi, N., Jalali, H., & Ghadiri, M. (2021). Lexical bundles in the abstract and conclusion sections: The case of applied linguistics and information technology. *Applied Research on English Language*, 10(3), 47-76.
- Shahriari Ahmadi, H., Ghonsooly, B., & Hosseini Fatemi, A. (2013). Analyzing research article introductions by Iranian and native English-speaking authors of Applied Linguistics. *International Journal of Research*, 2(3), 3-18.
- Shirazizadeh, M., & Amirfazlian, R. (2021). Lexical bundles in theses, articles and textbooks of applied linguistics: Investigating intradisciplinary uniformity and variation. *Journal of English for Academic Purposes*, 49, 100946.
- Shokri, N. M. (2010). Team project facilitates language learning. *Procedia-Social and Behavioral Sciences*, 7, 555-564.
- Sihotang, A. M., Sitanggang, F., Hasugian, N., & Saragih, E. (2021). The Effective Way to Develop Speaking Skills. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, 9(1).
- Sykes, D. L. (2017). *An investigation of spoken lexical bundles in interactive academic contexts* [Carleton University].
- Unsworth, L., & Mills, K. A. (2020). English language teaching of attitude and emotion



- in digital multimodal composition. *Journal of Second Language Writing*, 47, 100712.
- Valipoor, L. (2010). A corpus-based study of words and bundles in chemistry research articles (Unpublished Master's thesis). *Kashan University, Kashan*.
- Verzella, M. (2020). Hospitable writing: accommodating emergent users of English by means of intralingual translation. *Linguistics and Education*, 56, 100791.
- Vo, S. (2019). Use of lexical features in non-native academic writing. *Journal of Second Language Writing*, 44, 1-12.
- 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.-C. (2012). An exploration of vocabulary knowledge in English short talks: A corpus-driven approach. *International Journal of English Linguistics*, 2(4), 33.
- Wei, Y., & Lei, L. (2011). Lexical bundles in the academic writing of advanced Chinese EFL learners. *RELJ journal*, 42(2), 155-166.
- Wood, D. (2006). Uses and functions of formulaic sequences in second language speech: An exploration of the foundations of fluency. *Canadian Modern Language Review*, 63(1), 13-33.
- Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied linguistics*, 21(4), 463-489.
- Wray, A., & Fitzpatrick, T. (2008). Why can't you just leave it alone? Deviations from memorized language as. *Phraseology in foreign language learning and teaching*, 123.
- 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.
- 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*, 475-493.
- Yoon, C., & Choi, J.-M. (2015). Lexical bundles in Korean university students' EFL compositions: A comparative study of register and use. *Modern English Education*, 16(3), 47-69.
- Yousaf, M., & Shehzad, W. (2018). Prevalence of prefabricated structures in academic discourse: A corpus-based study. *International Journal of English Linguistics*, 8(5), 297-306.
- Zipagan, M. N., & Lee, K. R. (2018). Korean English Learners' Use of Lexical Bundles in Speaking. *Journal of Asia TEFL*, 15(2), 276.

### Biodata

**Ashraf Vaziri** is a Ph.D. candidate in TEFL at Islamic Azad University, Ayatollah Amoli Branch. She is interested in both teaching and translation. She has taught English at different levels of language proficiency and currently teaches translation courses and applied linguistics. She is also translated different articles and short stories. She has taught at high school for 7 years and enjoys working with teenagers.  
Email: [ashrafv294@yahoo.com](mailto:ashrafv294@yahoo.com)

**Hamed Barjesteh** is an associate professor of applied linguistics, and he is the head and a faculty member of Islamic Azad University, Ayatollah Amoli Branch, Amol, Iran. A teacher educator for 20 years, Hamed Barjesteh has published several books and articles on teaching English to young and adult language learners and presented at conferences at various national and international levels. He has trained teachers and supervised many MA theses and Ph.D. dissertations. He serves as the head of English Language department, an editor and a reviewer for various journals and conferences. His research interests lies in the field of critical language pedagogy, thinking skills in language learning, critical language testing, dynamic assessment, and corrective feedback.  
Email: [h.barjesteh@iauamol.ac.ir](mailto:h.barjesteh@iauamol.ac.ir)  
[h.barjesteh@iauamol.ac.ir](mailto:h.barjesteh@iauamol.ac.ir)

**Atefeh Nasrollahi Mouziraji** is an assistant professor in TEFL at Islamic Azad University, Ayatollah Amoli Branch. She has published a number of papers and participated in many national and international conferences. Her main areas of interests are first language acquisition, learning styles, literature teaching, critical thinking and pedagogy.

Email: [nasrollahi.atefeh59@gmail.com](mailto:nasrollahi.atefeh59@gmail.com)

