International Journal of Research Studies in Language Learning
2013 July, Volume 2 Number 3, 3-18

Analyzing research article introductions by Iranian and native English-speaking authors of Applied Linguistics

Shahriari Ahmadi, Hesamoddin
Ferdowsi University of Mashhad, Iran (hesamshahriari@yahoo.com)
Ghonsooly, Behzad
Ferdowsi University of Mashhad, Iran (ghonsooly@yahoo.com)
Hosseini Fatemi, Azar
Ferdowsi University of Mashhad, Iran (azar.h.fatemi@gmail.com)

ISSN: 2243-7754
Online ISSN: 2243-7762

OPEN ACCESS

Received: 31 July 2012 **Revised**: 17 August 2012 **Accepted**: 1 September 2012
Available Online: 7 September 2012 **DOI**: 10.5861/ijrsll.2012.158

Abstract

Considering the role of academic writing in the dissemination of research findings among members of scientific communities, efforts have been made by language specialists to achieve a more detailed understanding of this register, with hopes that the derived features could then be used to teach researchers, writing in a second/foreign language, to more effectively communicate the results of their studies. This study sets out to achieve a similar goal through the analysis of research article introductions in terms of the frequency and function of their lexical bundles. To this end, a corpus of 200 research article introductions by published authors in the field of Applied Linguistics was compared to a similar corpus by Iranian, non-native writers of the same field. The findings reveal that Iranian authors use more 4-word lexical bundles in their writing compared to their native-speaker counterparts. Syntactic and functional differences between the two corpora are introduced and implications for academic writing instructors are discussed in detail.

Keywords: lexical bundles; multiword expressions; corpus linguistics; research articles

Analyzing research article introductions by Iranian and native English-speaking authors of Applied Linguistics

1. Introduction

Academic writing registers, especially the research article, have become the subject of numerous ESP studies. This trend could be attributed to the value placed upon publishing research results and sharing one's findings with the academic community. In addition, in many countries, graduate students are also expected to publish their work in international journals, meaning they have to develop advanced writing skills in their own field of specialty. In Iran, English is the language of communication for some designated university majors; this particularly holds true for the field of English language studies. In Iranian universities, English is the predominant language used in textbooks, classroom lectures and discussions for students of English, in general, and Applied Linguistics in particular. University entrance exams include general language proficiency items, which guarantee the admission of those students who are more advanced in terms of their general English language proficiency. However, despite the advanced level of communicative skills in English, students are still found struggling when it comes to assignments involving academic writing. This comes as no surprise, since even native speakers require formal training in order to gain the ability to write academic research articles.

Assuming there is a continuum of proficiency in academic writing, where fully-competent writers of research articles stand at one extreme, it would be highly useful to locate where advanced learners are currently placed along the continuum and what are the gaps which need to be bridged for them to achieve more efficient writing skills. This is a delicate task, because differences between the writing of such advanced level students and the norms established within the register cannot be labeled as mistakes or errors, and should rather be referred to as deviations from the norm. Such deviations are very difficult to single out relying only on one's intuition. This is where corpus linguistics methods can be of substantial help to the researcher. Instead of relying on intuitions regarding the nature of language, the corpus-based approach to linguistic analysis draws heavily upon empirical data to answer language-related questions. The applications of corpus-based research are numerous, which is why an increasingly greater number of researchers are adopting corpus tools for answering their research questions. Besides their applications in materials development (Flowerdew, 1996), language corpora have been used to improve various aspects of writing such as knowledge of grammar (Diniz & Moran, 2005), genre awareness (Tribble, 2002) and vocabulary knowledge (Nation, 2001; Altenberg, 1998).

1.1 The corpus-driven approach to academic writing

Corpus-driven methodologies can be used to analyze and identify the most frequent or infrequent features within learner language (i.e., common grammatical structures, lexis, and discourse items), comparing them with the standard norms within the register, with the purpose of improving the quality of pedagogy and materials development. Most studies in this area have set their sights on describing the linguistic features of academic research articles. While some have focused on providing a thorough description of language used in a particular scientific field (e.g., Halliday, 1988), others have attempted to compare and contrast features of academic language with that of other genres (e.g., Biber, Johansson, Leech, Conrad, & Finegan, 1999). There have also been a series of studies directed towards investigating the various aspects of academic language as produced by learners. For instance, there have been several studies aimed at comparing academic language by foreign/second language learners with that of their native-speaker counterparts (Granger, 1998). In fact, a number of large-scale learner corpora have been compiled and are extensively used for the above-mentioned purposes. Leech (1998) maintains that a collection of learner texts representing learner language is a valuable resource to teachers and researchers wishing to gain a better understanding of how languages are learned and how the learning process can be improved.

Almost all of the established learner corpora in the field exclusively include written academic language. This could either be attributed to the importance of academic writing as a register or to the relative ease with which such corpora can be compiled, as compared to spoken registers. The most well-known learner corpora currently used in the field also vary depending on the background of the learners from whom texts are collected. While some corpora include learner texts from a variety of language backgrounds, others limit themselves to a particular learner group (e.g., Cantonese, Polish or Hungarian speakers). Each of these two categories of learner corpora can be used to provide answers to different types of research questions, depending on their specificity. Variations can also be seen in terms of the proficiency level of the target learners and the genres and styles included into the corpus (e.g., argumentative essays, exam texts, etc.). Some of the most widely-recognized learner corpora include the International Corpus of Learner English (ICLE), the Upsala Student English corpus (USE), the Cambridge Learner Corpus (CLC), and the Longman Learners' Corpus (LLC), just to name a few.

1.2 Expert writing vs. non-native writing

A corpus comparison between academic writing by advanced, non-native writers and professional authors writing within the same register can help define the existing differences between the two. These qualitative differences in language use can then serve multiple purposes. First, they can be used to create awareness among advanced learners regarding the existing gaps in their writing proficiency. In other words, it would serve to inform them of what changes could be made to their already-advanced writing skills in order to bring it closer to what is standard among the members of the academic writing discourse community. Second, the findings of such a study would be of great value to writing instructors who are preparing graduate students for writing publishable research articles. If such instruction were based on empirical evidence found in large corpora, in contrast to intuitive speculations regarding the norms of academic writing, it would most probably result in more successful writers and greater ease in publishing.

1.3 Lexical bundles

One linguistic unit which has become the subject of numerous comparative corpus studies in recent years is the lexical bundle. Biber and Conrad (1999) define lexical bundles as a sequence of words which recur with a relatively high frequency within a register. Lexical bundles are different from idioms in that they are non-compositional, semantically transparent and occur far more frequently in discourse. In addition, they are not always structurally complete. Lexical bundles are also commonly found in great numbers in academic registers. For example, although most words within a text do not occur in recurrent combinations, around 21% of the academic sub-corpus of the Longman corpus of Spoken and Written English appear in frequent bundles, some seen over 200 times per million words. When extracting lexical bundles from a corpus, a purely frequency-based approach is adopted (Altenberg, 1998; Butler, 1997; Biber et al., 1999). Therefore; the size of the bundle can affect the probability of its occurrence. For example, a 4-word bundle can be expected to be found less frequently than a 3-word bundle.

1.4 Categories of lexical bundles

Lexical bundles have been categorized both structurally and functionally. The most widely-used structural analysis of bundles has been proposed by Biber et al. (1999) who categorizes bundles into three types: independent verb fragments, verb fragments with dependent clause elements and noun phrase fragments. The first type includes verb fragments, beginning with a subject pronoun followed by a verb phrase. This type of bundle could also begin with a verb phrase, without a pronoun or question fragment. Bundles beginning with discourse markers followed by a verb phrase and question fragments are also categorized under this category. The second type of bundle characterized by verb phrase elements incorporating dependent clause fragments often include a complementizer followed by a main clause or a WH-word presenting a dependent clause. Dependent clause fragments starting with a complementizer or subordinator also form Type 2 bundles. The third categories mainly include noun phrase components ending in a post-modifier, prepositional phrase component

with modifiers or involve comparative expressions.

1.5 Previous comparisons between expert and non-native writing

Multiple studies have attempted to compare learner- and native-speaker corpora based on their frequency and type of lexical bundles. Cortes (2002) compared the use of lexical bundles by university students and published authors in the disciplines of history and biology. The results of her study revealed that students used fewer lexical bundles compared to professional writers, and there was very little overlap between the bundles used by the two groups. Baker and Chen (2010) made use of an automated frequency-driven method to draw a comparison between lexical bundles in published academic texts and student academic writing. This findings of this study also showed that learners use a comparatively narrower range of lexical bundles, but occasionally overuse specific lexical sequences that are rarely encountered in the corpus of texts by professional academic writers. In another study, Wei and Lei (2011) looked at lexical bundles across a corpus of doctoral dissertations by advanced Chinese EFL writers and published research articles by professional writers. Contrary to the previous two studies, the authors found that advanced learners used more frequent bundles and with a greater range. Generally, studies on lexical bundles have found differences between native and non-native language users both in terms of the frequency of bundles (Erman, 2009; Howarth, 1998; Adel & Erman, 2012) and in terms of the variety and range of bundles (Granger, 1998; Lewis, 2009).

1.6 Significance of the study

Previous studies in the area have looked at lexical bundles in research articles or dissertations as a whole. To our understanding, no study has looked into the use of lexical bundles in a particular sub-section of a register (except for the abstract, which is often seen as a separate genre in itself). The advantage of such an attempt is that through limiting the discourse functions of the texts, we can hope to extract more specific bundles, which consequently allow for more controlled comparisons. We also believe that the more limited range of functions in a specific section would lead to greater accuracy in the discoursal classification of bundles. In addition, the authors of the present study share the opinion that the highly technical nature of academic writing calls for a more detailed analysis. In other words, it would be simplistic to presuppose that all sections of the research article are similar in terms of their language features and style. A separate analysis of each section, however, would result in a more detailed understanding of the language used, and would consequently be of greater use to advanced learners. The present study compares Applied Linguistics research article introductions and literature reviews by professional writers for whom English is a native language and their Iranian non-native speaker counterparts with regards to the frequency and type of lexical bundles. The results of the analysis will hopefully inform the process of writing pedagogy as well as presenting a clear profile of academic writing in research article introductions in Applied Linguistics.

2. Method

2.1 The corpus

For this study, a 768,242-word corpus consisting of 400 research article introductions and literature reviews was compiled. From the 400 selected articles, 200 were authored by native speakers of English (NS), and published in accredited Applied Linguistics journals, namely *Applied Linguistics*, *TESOL Quarterly, English for Specific Purposes* and *Journal of Second Language Writing*. The remaining 200 articles were written by Iranian authors (IA) who had published their work in well-known Applied Linguistics journals inside Iran; these journals were *Journal of Teaching Language Skills, Iranian Journal of Applied Linguistics, Iranian Journal of Applied Language Studies* and *Research in Foreign Languages.* Each of the journals mentioned above contributed an equal number of articles (50) to the corpus. The NS sub-corpus consisted of 340,814 word tokens and 15,946 word types (The type-token ratio for this sub-corpus was 0.04), while the INNS corpus of introductions and

literature reviews consisted of 427,428 word tokens and 20,016 word types (The type-token ratio for this sub-corpus was also 0.04). All articles were randomly selected from different issues of the selected journals. Details of the compiled corpus have been presented in Table 1 below:

Table 1

Details of sub-corpora used in the analysis

Journals	(NS)	Journals	(INNS)
Applied Linguistics	50	Journal of Teaching Language Skills	50
TESOL Quarterly	50	Iranian Journal of Applied Linguistics	50
ESP Journal	50	Iranian Journal of Applied Language Studies	50
Journal of Second Language Writing	50	Research in Foreign Languages	50

2.2 Procedure

The corpus was searched for existing lexical bundles. In this study, we have adopted Biber et al.'s (1999) definition of lexical bundles, in which these units are defined as the most recurrent multi-word sequences within a particular register. This analysis was completed through a customized computer program written by the authors. The size of lexical bundles often varies across studies depending on the research purpose and other logistical elements such as the availability of units and their use in teaching and/or analysis. The present study focused on 4-word lexical bundles. The rationale behind this choice was that 4-word sequences appear more frequently than 3-word bundles and exhibit more clear structures compared to bi-grams or two word sequences (Hyland, 2008). Another decision in studies of lexical bundles is the frequency cut-off point for identification. In this study, sequences had to appear at least 5 times within the corpus to be considered as a recurring unit. Most studies tend to norm the counts so that the used corpora would be comparable. However, since the two corpora used in this study incorporated an equal number of texts (200), there was no need for norming. This decision ultimately leads to more identified bundles which are directly in line with the exploratory aims of the researchers. To avoid idiosyncratic expressions used by individual authors, another restriction was enforced, in which attestations for a given sequence would have to have occurred in at least 5 different texts.

Lexical bundles are known to not always represent complete structural units. Instead, they often serve to link two clausal or phrasal elements to each other. The relatively high frequency with which these bundles occur within a passage attests to their important role in language production and makes up for their lack of syntactic and semantic wholeness. In fact, by looking at lexical bundles, we can arrive at an index representing the degree to which the formulaic principle is prominent in the authors' use of language.

3. Results and discussion

The analysis of the two corpora resulted in a series of 4-word lexical bundles. The NS and INNS corpora contained 88 and 153 bundles, respectively. The most frequently recurring bundle in both corpora was *on the other hand*, which occurred 132 times in the INNS corpus and 66 times in the NS corpus. In general, among the ten most frequent bundles of each corpus, five were shared by both groups of writers. These included *on the other hand, in the field of, in the context of, the extent to which* and *at the same time*. The bundle *in the field of* ranked second in terms of frequency in the INNS corpus (with 64 attestations). The same bundle was the fifth most recurring bundle in the NS corpus, with 41 occurrences. The fourth most common bundle in each of the two corpora was *in the context of*, having been found 50 times in the INNS corpus and 46 times in the NS corpus. *The extent to which* was the third most common bundle used by native speakers (with 57 instances found in the NS corpus) and ranked sixth among the most recurring bundles in the INNS corpus (with 46 attestations). Finally, *at the same time* ranked ninth in both the NS and INNS corpora, with 37 and 40 attestations, respectively. Overall, 63 bundles were discovered to be shared by both groups of authors. These included bundles such as *English as a second, as a foreign language* and *of English as a*, which constitute chunks of characteristic phrases

in the field of Applied Linguistics (e.g., *of English as a second language*). Tables 2 and 3 below show the list of 4-word lexical bundles along with their frequency in the NS and INNS corpora, respectively. The lexical bundles common between the two corpora have been presented in bold.

Table 2

Most frequent 4-word lexical bundles found in the NS sub-corpus

Bundles	F	Bundles	F	Bundles	F
On the other hand	66	**In the form of**	20	**The present study is**	13
The use of English	59	In terms of their	20	Of English for Specific	12
The extent to which	57	**The use of the**	20	Use of English as	12
In the context of	46	The context of the	18	**An important role in**	12
In the field of	41	**In the process of**	18	**As a foreign language**	12
English for specific purposes	41	At the university of	18	From the perspective of	12
As well as the	37	A number of studies	18	**In the area of**	12
In the present study	37	**To be able to**	17	To the development of	12
At the same time	37	**In the target language**	17	Are more likely to	11
The nature of the	31	**In the United States**	17	**Extent to which the**	11
It is important to	31	Can be seen as	16	**Is based on the**	11
Native speakers of English	29	**Can be used to**	16	Of the English language	11
Of English as a	28	**As a result of**	15	**Studies have been conducted**	11
On the use of	27	**At the level of**	15	To refer to the	11
As a second language	27	Extent to which the	15	In the light of	11
In the case of	27	**Of this study is**	15	**It is necessary to**	11
Of the present study	26	To the use of	15	An understanding of the	10
The ways in which	25	**The role of the**	15	As a way of	10
That the use of	24	**English as a second**	14	**In a number of**	10
A wide range of	24	**In a variety of**	14	**In the use of**	10
In terms of the	24	**Number of studies have**	14	**Language teaching and learning**	10
On the one hand	23	**Through the use of**	14	Research in this area	10
Use of English in	22	**There has been a**	14	The development of the	10
The purpose of the	22	**The following research questions**	14	The meaning of the	10
The degree to which	21	**In the sense that**	13	**The relationship between the**	10
One of the most	21	**The importance of the**	13	**The results showed that**	10
On the basis of	21	**That there is a**	13	The use of a	10
English for academic purposes	21	In relation to the	13	**This study is to**	10
The results of the	21	The structure of the	13	Within the context of	10

The bundles extracted from the two corpora were classified both in terms of their structure and their function. The structural categorization of bundles in this study followed the taxonomy introduced by Biber et al. (1999). The majority of bundles found in both corpora were phrasal, which is typical of the academic genre. The NS and INNS corpora had 74 (83.1%) and 116 (75.8%) phrasal bundles, respectively. After phrasal bundles, verbal phrase fragments were the second most common type of bundles found in both groups of research article introductions and literature reviews. The NS corpus contained 11 (12.3%) such bundles and the INNS corpus included 29 (18.9%). Finally, the verb phrase element followed by a dependent clause was least found in both groups of texts. The NS and INNS corpora included 4 (4.4%) and 8 (5.2%) subordinate verb phrase elements, respectively. Figure 1 below illustrates the distribution of different types of bundles in the NS and INNS corpora.

Table 3

Most frequent 4-word lexical bundles found in the INNS sub-corpus

Bundles	F	Bundles	F	Bundles	F
On the other hand	132	**In the target language**	18	Of the fact that	13
In the field of	64	Investigated the effect of	18	**Studies have been conducted**	13
On the basis of	50	**Native speakers of English**	18	The development of a	13
In the context of	50	Of the most important	18	**Through the use of**	13
One of the most	49	**On the use of**	18	Learning a second language	13
The extent to which	46	**To be able to**	18	**The use of the**	13
As a foreign language	46	With respect to the	18	For the purpose of	13
As a result of	42	It is possible to	17	**It is important to**	13
At the same time	40	The present study was	17	A significant relationship between	12
English as a foreign	42	That there is no	17	Be considered as a	12
As well as the	38	**English as a second**	17	As a means of	12
Is one of the	36	In second language learning	17	**At the level of**	12
In the form of	36	**In the area of**	16	Can be used as	12
The following research questions	35	The fact that the	16	**Extent to which the**	12
On the one hand	33	Learning a foreign language	16	In a foreign language	12
In the process of	32	Is concerned with the	16	In a way that	12
In the case of	30	Of the use of	16	In the second language	12
On the part of	28	On the relationship between	16	**Is based on the**	12
The nature of the	28	A great deal of	16	Is referred to as	12
In the use of	27	**Can be used to**	15	Is there a significant	12
In the present study	26	**In a number of**	15	**Number of studies have**	12
That there is a	26	**In a variety of**	15	Of the study have	12
The degree to which	26	The basis of the	15	Of the study were	12
The results of the	24	Learning English as a	15	Should be noted that	12
A change of state	24	**The importance of the**	15	Significant difference between the	12
In terms of the	23	**This study is to**	15	**That the use of**	12
In terms of their	22	**It is necessary to**	15	The field of language	12
Of the target language	22	Is there any significant	15	To find out the	12
Of this study is	22	It should be noted	15	Of language teaching and	12
The present study is	22	**The ways in which**	15	Field of second language	12
To the fact that	22	Results of the study	15	**There has been a**	12
Of the present study	22	**The results showed that**	15	For the first time	11
The purpose of the	21	The field of second	15	Of Iranian EFL learners	11
Language teaching and learning	21	Investigated the relationship between	14	Of this study was	11
With regard to the	21	Language learning and teaching	14	On the role of	11
Of English as a	21	Answers to the following	14	**In the United States**	11
The relationship between the	21	At the end of	14	Studies have focused on	11
The role of the	21	On the nature of	14	The content of the	11
A wide range of	20	Purpose of this study	14	The last two decades	11
Is an attempt to	20	Studies have been carried	14	There is no significant	11
Second or foreign language	20	Study is an attempt	14	Were found to be	11
The Iranian EFL learners	20	The end of the	14	Important role in the	10

Table 3 … continue

Most frequent 4-word lexical bundles found in the INNS sub-corpus

Bundles	F	Bundles	F	Bundles	F
A large number of	20	The language of the	14	Degree to which the	10
An important role in	19	The part of the	14	Findings of this study	10
Beliefs about language learning	19	The process of learning	14	**In the sense that**	10
Have been carried out	19	**As a second language**	14	Is likely to be	10
Is there any relationship	19	In the study of	14	Is related to the	10
In an attempt to	19	The results indicated that	13	Of the nature of	10
The purpose of this	19	And the use of	13	The relationship between language	10
Of teaching and learning	19	Can be considered as	13	This study was to	10
There any relationship between	19	In a second language	13		
As one of the	18	In second language acquisition	13		

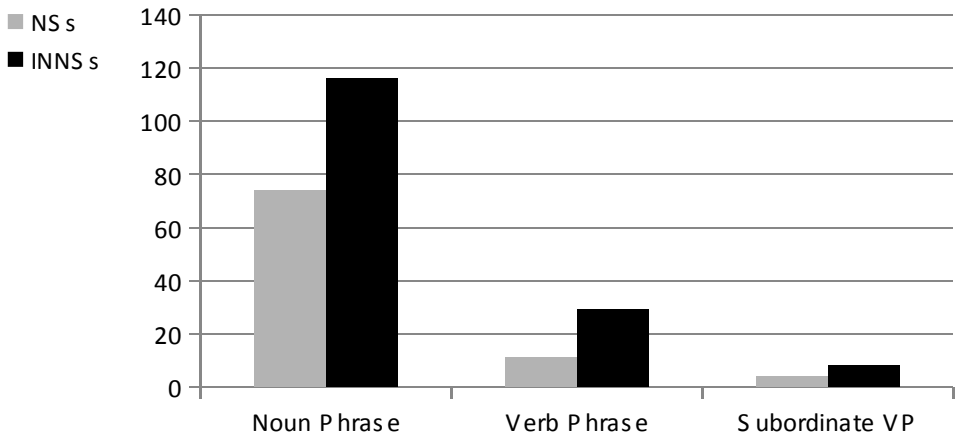

Figure 1. Distribution of noun phrase, verbal phrase and subordinate verbal phrase bundles in the NS and INNS corpora

A functional analysis involves the classification of lexical bundles into distinct categories based on the discourse purposes they serve in texts. The present study followed a top-down approach to the analysis of discourse functions (Biber, Connor, & Upton, 2007). Through such an approach, identified bundles are categorized into distinct discourse categories based on a specific analytical framework. Two analytical frameworks were used in this study. One was the taxonomy developed by Cortes (2002) and later improved by Biber et al. (2007). The second was the revised genre prototype developed by Swales (2004).

In the first framework, three major functional types are introduced: *Stance Bundles, Discourse Organizers* and R*eferential Expressions. Stance Bundles* refer to that group of bundles which reveal the attitude of the writer or judgments regarding the degree of certainty, proposition or ability. Bundles belonging to the group of *Discourse Organizers* serve to structure and organize the text by clarifying or elaborating on the topic. The third, category of lexical bundles, *Referential Expressions*, reference physical or abstract entities or the textual context itself. It is possible for a specific lexical bundle to express different functions. The table below provides a more detailed description of the taxonomy used in this study.

The lexical bundles were initially categorized based on the first analytical framework. As expected from written academic registers, the most prevalent type of bundle in both corpora was Referential Expressions (with 73 and 119 instances in the NS and INNS corpora, respectively). Discourse Organizers were the second most common type of bundle (with 8 and 24 instances in the NS and INNS corpora, respectively). Finally, the least

common functional category of lexical bundles, based on the first framework, was Stance Expressions (with 7 and 10 instances in the NS and INNS corpora, respectively). Table 5 shows the number and percentage of different types of lexical bundles classified according to the first analytical framework.

Table 4

Functional classification of lexical bundles

Lexical Bundle Category	Example
I. Stance Expressions	
a. Epistemic Stance	
Personal	*I don't know if*
Impersonal	*Are more likely to*
b. Attitudinal/Modality Stance	
b.1. Desire	*I don't want to*
b.2. Obligation/Directive	
Personal	*You might want to*
Impersonal	*It is important to*
b.3. Intention/Prediction	
Personal	*What we're going to*
Impersonal	*Are going to be*
b.4. Ability	
Personal	*To be able to*
Impersonal	*Can be used to*
II. Discourse Organizers	
a. Topic Introduction/Focus	*Let's have a look*
b. Topic Elaboration/Clarification	*On the other hand*
III. Referential Expressions	
a. Identification/ Focus	*One of the most*
b. Imprecision	*Or something like that*
c. Specification of Attributes	
c.1. Quantity Specification	*The rest of the*
c.2. Tangible Framing Attribute	*In the form of*
c.3. Intangible Framing Attribute	*The nature of the*
d. Time/Place/Text Reference	
d.1. Place Reference	*In the United States*
d.2. Time Reference	*At the time of*
d.3. Text Deixis	*As shown in figure*
e. Research Reference Bundles	*The following research questions*
f. Subject-Specific Bundles	*English for Specific Purposes*

According to the second discourse analytical framework, research article introductions as a genre are divided into three major moves: *Establishing a territory, establishing a niche* and *presenting the present work.* Each of these three steps, in turn, consists of a number of moves. An outline of the proposed moves and steps is presented in table 5 below.

Table 5

Swales' revised genre prototype for research article introductions (2004, pp. 230, 232)

Move 1:	Establishing a territory (citations required) via Topic generalizations of increasing specificity
Move 2:	Establishing a niche (citations required) via:
	Step 1A: indicating a gap, or
	Step 1B: Adding to what is known
	Step 2: Presenting positive justification (optional)
Move 3	Presenting the present work via:
	Step 1: Announcing the present research descriptively and/or purposively (obligatory)
	Step 2: Presenting research questions or hypotheses* (optional)
	Step 3: Definitional clarifications* (optional)
	Step 4: Summarizing methods* (optional)
	Step 5: Announcing principal outcomes (optional)**
	Step 6: Stating the value of the present research (optional)**
	Step 7: Outlining the structure of the paper (optional)**

Notes: * Steps 2-4 are less fixed in their order of occurrence than the others. ** Steps 5-7 are probable in some fields, but unlikely in others.

Table 6

Number and Percentage of lexical bundles according to the first analytical framework

Lexical Bundle Category	NS - Number	NS - Percentage	INNS-Number	INNS (%)
I. Stance Expressions				
a. Epistemic Stance				
Personal	0	0.00%	0	0.00%
Impersonal	0	0.00%	1	0.65%
b. Attitudinal/Modality Stance				
b.1. Desire	0	0.00%	0	0.00%
b.2. Obligation/Directive				
Personal	0	0.00%	0	0.00%
Impersonal	3	3.40%	7	4.57%
b.3. Intention/Prediction				
Personal	0	0.00%	0	0.00%
Impersonal	1	1.13%	0	0.00%
b.4. Ability				
Personal	0	0.00%	0	0.00%
Impersonal	3	3.40%	2	1.30%
II. Discourse Organizers				
a. Topic Introduction/Focus	4	4.54%	7	4.57%
b. Topic Elaboration/Clarification	4	4.54%	17	11.11%

Table 6 … continue

Number and Percentage of lexical bundles according to the first analytical framework

Lexical Bundle Category	NS - Number	NS - Percentage	INNS-Number	INNS (%)
III. Referential Expressions				
a. Identification/ Focus	3	3.40%	5	**3.26%**
b. Imprecision	0	0.00%	0	0.00%
c. Specification of Attributes				
c.1. Quantity Specification	2	2.27%	4	2.61%
c.2. Tangible Framing Attribute	0	0.00%	0	0.00%
c.3. Intangible Framing Attribute	44	50.00%	58	37.90%
d. Time/Place/Text Reference				
d.1. Place Reference	4	4.54%	1	0.65%
d.2. Time Reference	1	1.13%	3	1.92%
d.3. Text Deixis	0	0.00%	0	0.00%
e. Research Reference Bundles	7	7.95%	18	11.76%
f. Subject-Specific Bundles	12	13.63%	30	19.60%

As can be seen in Table 5, considering the percentage of bundles related to each of the three functions, it appears that native-speakers made use of a greater proportion of lexical bundles for describing their attitudes and assessments and also reflecting the relationship between prior and coming discourse. On the other hand, Iranian non-native speakers used a larger fraction of their bundles for referencing purposes. It is also interesting to note that Iranian authors used relatively more subject-related and research-referencing bundles in their writing. Subject-related bundles refer to those 4-word sequences that referenced a concept or notion specific to the field of Applied Linguistics. Some examples of such bundles are:

... English by Brazilians and in English by native speakers of English (NS Corpus)

... that deserve investigation in **language teaching and learning** circles are fear of... (INNS Corpus)

On the other hand, research-related bundles were those expressions which made reference to the study itself or other reports of research. These bundles were sometimes difficult to identify, because by just looking at the lexical bundle itself, one could not always tell whether the expression was referencing a research study or any of its elements. Therefore, the KWIC (Key Word in Context) tool was used whenever references were not clear. Below are two examples of research-related expressions from the NS and INNS corpora.

The value **of the present study** lies in providing better information about (NS Corpus)

To achieve this purpose, **the following research questions** were addressed (INNS Corpus)

The decision to include a second framework in the functional analysis of lexical bundles came from the need to classify bundles according to genre-specific criteria. That is, the functional taxonomy provided by Biber et al. (2007) is a general one that can be used to classify lexical bundles coming from texts of different registers. By using Swales' (2004) genre prototype for article introductions, we hope to achieve a more genre-specific classification that can be of greater use to members within the specific discourse community. However, there are two major problems with using such a framework. First of all, in a corpus-based move analysis, researchers must first identify and tag the moves and move types in each individual text (Biber et al., 2007). This is a time-consuming and labor-intensive task and considering the relatively large sample of texts used in this study, would have been extremely difficult. Second, not all of the bundles found in the analysis can be easily classified

according to moves and steps. Although texts of the same register exhibit similarities in terms of discourse objectives, there is a large degree of variation among them both in terms of content and linguistic characteristics. Therefore, to expect all bundles to neatly fit into a particular move or step category would be over-simplifying matters. As a result, the two frameworks of analysis used in this study are meant to have a complementary relationship with one another and provide better insight into the functions of lexical bundles through providing different perspectives.

During the classification process, the two raters were told to initially rely on their intuitions as highly proficient users of English, to determine whether a bundle served the purpose of any of the three major moves (*establishing a territory, establishing a niche* and *presenting the present work*). Following this step, the raters were required to validate their judgment by referring to the KWIC pertaining to the bundle in question. If more than one-third of the attestations contained the lexical bundle with a specific function, that bundle would be classified as fulfilling the observed move. For example, if 4 attestations out of 12 included a lexical bundle used to present the current study, it would be inferred that the lexical bundle was used to fulfill the third move in Swales' model. However, because the corpus was not coded for steps prior to the analysis, it was difficult to determine specifically which steps were being fulfilled by bundles in addition to specifying their move category. Some of the bundles were more easily classified relying on intuition, while others were more difficult and required the raters to read through the attestations carefully. For example, the bundle *in the present study* could easily be classified as presenting the present work (Move 3). On the other hand, the bundle *is one of the* cannot be easily classified without referring to context. However, by looking at the KWICs below, it becomes clear that this bundle is used to establish a territory for the readers.

accent **is one of the** first noticeable features of oral communication, (INNS Corpus)

Gender **is one of the** factors that can inherently be of interest to... (INNS Corpus)

vocabulary learning **is one of the** most problematic areas of language learning... (INNS Corpus)

self-confidence **is one of the** most important determinants in learner motivation... (INNS Corpus)

At the opposite end of the continuum, some bundles, such as *the structure of the*, seem to be obvious in their relation to a specific move (in this case Move 3); but once we look at the contexts in which these sequences occur, we soon realize that our intuitions have been misleading. In this case, *the structure of the* is not used to outline the structure of the paper. Rather, as in the case of many other bundles, it cannot be classified into any of the three move categories based on the sentences in which it occurs.

... understanding of **the structure of the** hypertext have fewer difficulties navigating ... (NS Corpus)

... a "fundamental relation" between "**the structure of the** body of knowledge of a given discipline and ... (NS Corpus)

Some bundles were seen to fulfill different functions at the same time. For instance, the sequence *there has been a* was used both for establishing a territory (Move 1) and establishing a niche (Move 2). The first sentence below shows the use of this bundle for Move 1 and the second sentence demonstrates how the exact same bundle can be used in Move 2.

In recent years **there has been a** significant growth in the literature on the role of ... (NS Corpus)

Despite these recent studies, **there has been a** general lack of inquiry on how L2 writers perceive ...(NS Corpus)

Even more interesting are those bundles which relate to different moves, depending on the corpus they appear in. For example, the bundle *it is important to* serves to establish a territory (Move 1) in the INNS Corpus

and in the NS corpus, it is mainly used to establish a niche (Move 2). The examples below clearly show how this bundle is used by native and non-native writers to different ends.

It is important to point out that the only test which has followed this procedure to date is... (INNS Corpus)

... programs on ESL student academic achievement, **it is important to** study not only short-term student success rates ... (NS Corpus)

The analysis revealed that in the NS corpus, from the total number of identified lexical bundles (88), 52 could be classified according to the second analytical framework (59.09%). As for the INNS corpus, 91 bundles (from 153) could be categorized based on Swales' model (59.47%). In both corpora, the most number of bundles were used for the purposes of Move 1 (37 and 47 bundles in the NS and INNS corpora, respectively). Following Move 1 expressions, native speakers made greater use of bundles aimed at describing the niche (Move 2), with 10 bundles. However, non-native speakers were different in that they used Move 3 bundles (26) more commonly than Move 3 (18). Table 6 summarizes the different functional types of lexical bundles based on the second analytical framework.

Table 7

The number and percentage of different types of lexical bundles according to the second analytical framework

	NS - frequency	NS - Percentage	INNS - frequency	INNS - Percentage
Move 1 Bundles	37	71.15%	47	51.60%
Move 2 Bundles	10	19.23%	18	19.78%
Move 3 Bundles	7	13.46%	26	28.57%

As can be seen in Table 6, in both corpora, Move 1 bundles were more frequent than the other categories; but the percentages reveal that native speakers used a greater proportion of their lexical bundles to define the area of research. These bundles mostly include subject-specific bundles (e.g., *as a second language, English for Academic Purposes*) and intangible framing attributes (e.g., *in the field of, the nature of the*). Both groups of writers used almost the same rate of Move 2 bundles. Move 2 bundles were mostly of a verbal phrase (e.g., *studies have been conducted, studies have focused on*) and subordinate verbal phrase structure (e.g., *the results showed that, it should be noted*). Another characteristic of Move 2 bundles was that they were either referential expressions (*the results of the*, *research in this area*) or topic clarification discourse organizers *(have been carried out, it is necessary to*).

The difference in the total number of lexical bundles found within the two corpora and the fact that non-native writers in this study used more bundles corroborates the findings of other studies that have also reported greater use of lexical bundles by advanced EFL/ESL writers (Cortes, 2004; Hyland, 2008; Pang, 2009 and Wei and Lei, 2011). Different studies have attempted to investigate the reasons behind the advance non-native speakers' tendency to use lexical bundles more frequently and repetitively. Some of these studies have attributed this characteristic of non-native writing to the perception of prefabricated forms and patterns as reliable safety nets which can be used confidently, especially at times of uncertainty (De Cock, 2000; Granger, 1998). That is, the use of lexical bundles can be seen as a form of compensation strategy by non-native writers of English, trying to ensure the correctness and appropriateness of the language they use. In Granger's (1998) own words, non-native writers exhibit a tendency towards clinging on to "certain fixed phrases and expressions which [they] feel confident in using (p. 156).

The present study analyzed the identified lexical bundles according to two analytical frameworks. The main advantage of such an approach to the functional analysis of lexical bundles is that it provides a fresh perspective towards the categorization of these linguistic features, while retaining the previously-established taxonomy for comparative purposes. The first analysis found that referential bundles were the most common category in both

corpora. This finding is in line with the general view that referential bundles are one of the most commonly found bundles in written academic registers. Biber and Baribieri (2007) analyzed four academic registers in their study and found that in three of the registers, referential bundles were the most common type. Also, in both groups stance expressions were not seen as frequently as other the other two functional categories. These stance expressions are mostly used as hedging devices, presenting the authors' assessment regarding the certainty of claims and propositions. The less frequent use of such bundles shows that as pointed out by previous researchers (Hyland, 1994; Lorenz, 1999), academic writing, especially L2 academic writing, shows more limited control on cautious language. However, the difference between the two groups of texts with regard to their proportional use of such bundles was not great, meaning that both groups generally made more limited use of stance bundles.

The syntactic classification of lexical bundles based on Biber et al.'s (1999) taxonomy showed that Iranian writers used considerably more verb phrase fragments and verb phrase elements followed by dependent clauses. Biber, Gray, and Ponpoon (2011) maintain that clausal subordination is more commonly encountered in academic conversation compared to academic writing, which generally tends to include more noun phrase constituents and complex phrasal structures. Considering these findings, Iranian EFL writers, in spite of their high level of language proficiency and academic writing expertise as specialists in the field of Applied Linguistics display a tendency towards writing in a conversational style, relying heavily on clausal elements and subordination. This outcome is not unexpected, because academic writing with its dependence on extended noun phrases and limited use of verbs is different from most other registers and can only be seen in academic writing, which aims to achieve greater brevity and a more density, conveying more information in fewer words.

4. Conclusion

This study investigated the use of 4-word lexical bundles in research article introductions by native English speakers and Iranian EFL writers in the field of Applied Linguistics. Through a corpus-based approach to linguistic analysis, the two groups of introductions were compared in terms of the number, type and function of lexical bundles. Results revealed that advanced writers tend to use more multi-word expressions in their writing. The lexical bundles were functionally categorized based on two frameworks of analysis. The results of the analysis according to the first framework showed greater conformity between the two groups of texts (Referential expressions were the most common type followed by discourse organizers and stance expressions). However, the results of the analysis based on the second analytical framework revealed that following the use of lexical bundles for establishing the area or realm of study (Move 1), non-native authors mostly used the second most number of bundles for referencing their research; while non-native writers used the second most number of bundles for establishing a niche.

Writing instructors can make use of the findings of this study by further emphasizing the role of lexical bundles for their students, explaining to them the various ways through which these prefabricated language patterns can be used to achieve moves and steps. This would consequently help them achieve writing that is more understandable to target readers and heightens their chances of publishing their work in accredited journals. Future studies can identify more lexical bundles used in different academic disciplines and construct an inventory of lexical bundles, used as genre frames. Such an inventory would most certainly be of great interest to academic writing instructors and advanced EFL writers alike.

The application of two different discourse frameworks in the functional analysis of lexical bundles in this study is also of interest. The fact that the results of one analysis led to the identification of differences between the two groups of texts, while results from the second analysis did not, shows that the selection of frameworks is of great methodological importance in lexical bundle studies. These frameworks are like lenses through which researchers can observe their subject of interest. Each of these lenses provides a different, yet equally noteworthy, view of the subject of analysis. That is, depending on what framework researchers choose for their investigations, they can expect different outcomes with different implications. For example, in this study, results from the first functional analysis showed that both groups of writers make infrequent use of stance expressions and hedging

devices. However, in the second analysis, we can see that native speakers use a larger percentage of their lexical bundles to introduce the realm of their study, compared to their non-native counterparts. One the other hand, non-native writers use a much greater percentage of their bundles to refer to their own study.

The results of this study also have implications for writing instructors in the field of English for Academic Purposes (EAP). According to frequency models of language learning, phraseological units are learned probabilistically and as a result of strengthening associations among co-occurring words (Ellis, 2002). The more frequently a learner is exposed to particular bundles, the higher the chances are for that learner to produce the bundles in speech or writing. This linear relationship between exposure and output is often times affected by L1 background. If a bundle shares an equivalent in the L1, the probability of its use will increase. As previously discussed, the proclivity towards extended noun phrase elements in the academic writing register is unnatural in that it is rarely seen in other registers. It is also very uncommon for other languages (in this case, Farsi) to follow such a trend; hence, learners, even those at advanced proficiency levels, find it counterintuitive and avoid using bundles representing noun phrase elements. As a result, it is recommended that EAP teachers and academic writing instructors create necessary awareness and improve the quality of academic writing by exposing learners to high-frequency phrasal bundles used by native speakers.

5. References:

Adel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundle approach. *English for Specific Purposes, 31*, 81-92. <http://dx.doi.org/10.1016/j.esp.2011.08.004>

Altenberg, B. (1998). On the phraseology of spoken English: the evidence of recurrent word combinations. In A. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp.101-122). Oxford: Oxford University Press.

Butler, C. S. (1997). Repeated word combinations in spoken and written text: Some implications for Functional Grammar. In C. S. Butler, J. H. Connolly, R. A. Gatward, & R. M. Vismans (Eds.), *A fund of ideas: recent developments in functional grammar* (pp. 60-77). Amsterdam: IFOTT, University of Amsterdam.

Biber, D., & Baribieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263-286. <http://dx.doi.org/10.1016/j.esp.2006.08.003>

Biber, D., Gray, B., & Ponpoon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45*(1), 5-35. <http://dx.doi.org/10.5054/tq.2011.244483>

Biber, D., Connor, U., & Upton, T. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure.* Philadelphia: John Benjamins.

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora: Studies in honor of Stig Johansson* (pp. 181-189). Amsterdam: Rodopi.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English.* London: Longman.

Baker, P., & Chen, Y. (2010). Lexical bundles in L1 and L2 academic writing. *Language, learning and technology, 14*(2), 30-49.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23,* 397-423. <http://dx.doi.org/10.1016/j.esp.2003.12.001>

Cortes, V. (2002). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23(4),* 397-423. <http://dx.doi.org/10.1016/j.esp.2003.12.001>

De Cock, 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.

Diniz, L., & Moran, K. (2005). Corpus-based tools for efficient writing instruction. *Essential Teacher, 2*(3), 36-39.

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

Flowerdew, L. (1996). Concordancing in language learning. In M. C. Pennington (Ed.), *The power of CALL* (pp. 97-114). Houston, TX: Athelstan.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: An interdisciplinary perspective* (pp. 145-160). Amsterdam: John Benjamins.

Halliday, M. A. K. (1988). On the language of physical science. In M. Ghadessy (Ed.), *Registers of written English* (pp. 162-178). London: Pinter.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics, 19*(1), 24-44. <http://dx.doi.org/10.1093/applin/19.1.24>

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

Leech, G. (1998). Preface. In S. Granger (ed.), *Learner English on computer* (pp. Xiv-xx). London: Longman.

Lewis, M. (2009). *The idiom principle in L2 English: Assessing elusive formulaic sequences as indicators of idiomaticity, fluency, and proficiency*. Saarbrucken, Germany: VDM Verlag.

Nation, I. S. P. (2001). Using small corpora to investigate learner needs: two vocabulary research tools. In M. Ghadessy, A. Henry and R. L. Roseberry (Eds.) *Small corpus studies and ELT* (pp. 31-45). Amsterdam: John Benjamins.

Pang, 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. *Teaching English in China, 32,* 25-45.

Swales, J. (2004). *Research genres: Explorations and applications.* Cambridge: CUP.

Tribble, C. (2002). Corpora and corpus analysis: New windows on academic writing. In J. Flowerdew (Ed.), *Academic discourse* (pp.131-149). London: Longman.

Wei, Y., & Lei, L. (2011). Lexical bundles in the academic writing of advanced Chinese EFL learners. *RELC Journal, 42(2)*, 155-166. <http://dx.doi.org/10.1177/0033688211407295>