# A Corpus-Based Comparative Study of Lexical Bundles in PhD Dissertation Abstracts in Physical Sciences and Arts and Humanities

ZHOU Bin[a],*

[a]Phd candidate in social-linguistics, Guangdong University of Foreign Studies, Guangzhou, China.
* Corresponding author.

## Abstract
This study aims to analyze Phd theses abstracts in terms of the frequency and function of lexical bundles. A corpus of of the PhD thesis abstracts collections in FLAX in the field of Physical Sciences and Arts and Humanities were compared. A linguistic analysis of the bundles shows that the lexical bundles found in the sub-corpus of Physical Sciences included more clausal elements and subordination, while bundles used in PhD thesis abstracts of Arts and Humanities were more phrasal in nature.

**Key words:** Lexical bundle; Multi-word expressions; Abstracts; Corpus linguistics

## INTRODUCTION

The scientific community has started to become increasingly globalized with English serving as the lingua franca. Researchers not only from English speaking countries but also from other language backgrounds have been and will be contributing to research in all fields of science. International journals accept publications from countries all around the world and this adds a certain amount of diversity to the use of English in research publications. Although the language of research is commonly thought of as being uniform, it has been shown through research that there is a certain amount of variation across disciplines and writers of different language and cultural backgrounds within academic language. Several language specialists including Brodkey (1987) and Crowley (1991) have even rejected the term "academic prose" due to the large-scale variability among texts in this register.

Researchers are becoming increasingly interested in variation within academic prose caused by various factors including language background and disciplines. Corpus linguistics is one of the methods commonly used in recent decades for exploring variation in language use owing to the practicality and reliability it offers in handling vast amounts of data. Corpus linguistics makes possible quantitative analyses of large amounts of language data with the use of computational analytical tools which would not have been possible through traditional methods of research. Studies contrasting specific corpora have proven the existence of variation even within a single genre e.g. among textbooks and research articles in biology and history (Conrad, 2001) among medical research articles ( Biber and Finegan, 2001) , among technical texts (Carrio-Pastor, 2009). As can be seen, the possibilities corpus linguistics provides for researching variation are endless.

Research on the lexical characteristics of academic discourse employing a corpus approach has also started to focus on specific subsets of language and various sections of research articles providing evidence for disciplinary and cross-cultural variation. Drawing from the idea that meaning is expressed differently in different text types, and Romer's (2010) suggestion to focus on subsets of language when studying phraseology, this study compares multi-word sequences in a corpus of Phd theses abstracts from a variety of disciplines, focusing on five-word bundles. Through this study, I intend to shed light on the differences between PhD theses writers from two academic backgrounds: Physical Sciences and Arts and Humanitie in regards to the most commonly used multi-word sequences in abstracts and their functions.

Based on the motivation of the present study, two research questions are listed as follows:

• What are the most frequent lexical bundles in the corpus of Phd theses abstracts from a variety of disciplines, focusing on 5-word bundles?

• What is the function of these lexical bundles in Phd theses abstracts?

# 1. LITERATURE REVIEW

The control of multi-word expressions is an important element of fluent language production and presents a challenge specifically to student writers and non-native speakers alike. For example, Romer (2010) who created a model for identifying the phraseological profile of a specific text type for investigating book reviews highlights the importance of knowledge about modification of common multi-word expressions, their functions and positions in a text for non-native writers. Academic clusters as central elements of academic discourse have been proven to provide important insights into the structure of academic discourse. Hyland's (2008) study on academic clusters in research articles, doctoral dissertations and master's theses has shown that research articles contain far fewer clusters which are mostly text-oriented in contrast to master's theses which display different patterns. Thus, academic clusters can function in differentiating genres and reflect the writers characteristics. For example Hyland (2008) interprets the high reliance on formulaic expressions by master's theses writers as a strategy employed by less confident or less proficient students in constructing their texts. The study of clusters in academic discourse would benefit instructional practices especially in advanced settings by allowing teachers to focus on the specific ways of creating meaning appropriate to particular kinds of writing rather than relying on the massive literature describing the research article (Hyland, 2008, p.60).

Existing studies on research article abstracts have revealed that although the format of abstracts change from journal to journal and from discipline to discipline (McCain, 1991), the language used in them is quite formulaic following a relatively restricted range of lexico-grammatical patterns (Hyland and Tse, 2004). For example twenty year analysis of abstracts by Cronin, Shaw and La Barre (2003) has shown that this established practice in research article publications reflects "symbolic and substantive significance enormously dependent on context, moment and behavioral characteristics of different scholarly tribes." (2003, p.109). Most studies on abstracts (Cronin and Franks, 2006; Cronin, Shaw & La Barre, 2003, 2004; Giles and Councill, 2004), however focus more on the social and intellectual role of abstracts and collaboration in the field of scholarly research but not the lexical characteristics and they do not follow a corpus linguistics approach.

# 2. METHOD

## 2.1 Corpus

**Table 1**
**Details of Sub-Corpora Used in the Analysis**

| Physical sciences | Arts and humanities |
| --- | --- |
| 000 Computer science, knowledge and systems | 050 Magazines, journals and serials |
| | 060 Associations, organisations and museums |
| | 070 News media, journalism and publishing |
| 500 Science | 100 Philosophy and psychology |
| 510 Mathematics | 110 Metaphysics |
| 520 Astronomy | 120 Epistomology |
| 530 Physics | 130 Parapsychology and occultism |
| 540 Chemistry | 140 Philosophical schools of thought |
| 550 Earth sciences and geology | 150 Psychology |
| | 160 Philosophical logic |
| | 170 Ethics |
| 620 Engineering | 180 Ancient, medieval and Eastern philosophy |
| 660 Chemical engineering | 190 Modern Western philosophy (19th-century, 20th-century) |
| 670 Manufacturing | 200 Religion |
| 680 Manufacture for specific purposes | 210 Philosophy and theory of religion |
| | 220 The Bible |
| | 230 Christianity |
| | 240 Christian practice and observance |
| 690 Construction of buildings | 250 Christian orders and local church |
| | 260 Social and ecclesiastical theology |
| 710 Area planning and landscape architecture | 270 History of Christianity |
| | 280 Christian denominations |
| 720 Architecture | 290 Other religions |
| | 400 Language |
| | 410 Linguistics |
| | 420 English and Old English languages |
| | 430 German and related languages |
| | 440 French and related languages |
| | 450 Italian, Romanian, and related languages |
| | 460 Spanish, Portuguese, Galician |
| | 470 Latin and Italic languages |
| | 480 Classical and modern Greek languages |
| | 490 Other languages |
| | 700 Arts |
| | 730 Sculpture, ceramics, and metalwork |
| | 740 Graphic arts and decorative arts |
| | 750 Painting |
| | 760 Printmaking and prints |
| | 770 Photography, computer art, film, video |
| | 780 Music |
| | 790 Sports, games and entertainment |
| | 800 Literature, rhetoric and criticism |
| | 810 American literature in English |
| | 820 English and Old English literatures |
| | 830 German and related literatures |
| | 840 French and related literatures |
| | 850 Italian, Romanian, and related literatures |
| | 860 Spanish, Portuguese, Galician literatures |
| | 870 Latin and Italic literatures |
| | 880 Classical and modern Greek literatures |
| | 890 Other literatures |
| | 900 History |
| | 910 Geography and travel |
| | 920 Biography and genealogy |
| | 930 History of ancient world (to ca. 499) |
| | 940 History of Europe |
| | 950 History of Asia |
| | 960 History of Africa |
| | 970 History of North America |
| | 980 History of South America |
| | 990 History of other areas |

The corpus used for the study is PhD Abstracts Collections; a sub-collection of the PhD thesis abstracts collections in FLAX (http://flax.nzdl.org/greenstone3/flax)

A Corpus-Based Comparative Study of Lexical Bundles in
PhD Dissertation Abstracts in Physical Sciences and Arts and
Humanities

and is derived from the EThOS Open Access toolkit, which is made available by the British Library for reuse by third parties for not-for-profit purposes. The copyright remains with the authors of the PhD theses unless otherwise stated in the licensing information provided by EThOS.

EThOS at the British Library uses the Dewey decimal classification system, which organizes library materials by discipline or field of study. The main divisions of the Dewey decimal classification system include philosophy, social sciences, science, technology, and history. These main divisions fall within a scheme made up of ten classes, each divided into ten divisions, and each having ten sections. Here with the PhD Abstracts Collections in FLAX, abstracts are directed into one of four over-arching discipline divisions to make up the following sub-collections: Arts and Humanities, Social Sciences, Physical Sciences and Life Sciences. This Physical Sciences sub-collection in FLAX includes abstracts that have been classified into discipline sub-divisions from the 000, 500, 600 and 700 classes of the Dewey decimal classification system, and the Arts and Humanities sub-collection from the 000, 100, 200, 400, 700, 800 and 900 classes of the Dewey decimal classification system, specifically.

### 2.2 Procedure

The adopted definition of lexical bundles is borrowed from Biber et al. (1999), who define them as the most recurrent multi-word sequences in a given register. The study focused on units of 5-word sequences. The frequency cut-off used for the identification of bundles was set at 4 times per million words. In comparison to other studies, this criteria is quite lenient, but given the investigative nature of our study and also, considering the short length of each text,

it best serves our purposes and can be more informative. This restriction also helps prevent methodological issues caused by the difference in the number of words in each corpus. In order to avoid idiosyncratic expressions by an individual author, it was agreed that a sequence would have to appear in at least 4 different texts to be included into the analysis. This decision was not restrictive, since the texts in the acknowledgement were too short to include recurrent expressions.

Comparisons were made between the sub-corpora in order to identify differences between PhD theses writers from two academic backgrounds: Physical Sciences and Arts and Humanitie in regards to the most commonly used multi-word sequences in abstracts and their functions.

## 3. RESULTS AND DISCUSSION

### 3.1 Extracted Lexical Bundles

The first round of analysis used for analysis in this study was to compare lexical bundles at the beginning of the sentence in two sub corpora. For example, *the aim of this thesis is to cast new light on such questions by approaching them from within the cognitive framework of Sperber and Wilson's Relevance Theory. The aim of this thesis* is a typical lexical bundle at the beginning of the sentence. 529 lexical bundles are retrieved in the sub corpus of Arts and Humanities, and 884 bundles in Physical Sciences corpus. The size difference of the two sub corpus results in the gap between their numbers of lexical bundles. Therefore, the 20 most frequent bundles found within the two corpora are picked out for further analysis, as shown in Table 2.

**Table 2**
**20 Most Frequent Bundles at the Beginning of Sentences within the Two Corpora**

| Physical sciences corpus | Frequency | Arts and humanities | Frequency |
|---|---|---|---|
| This thesis is concerned with | 124 | The aim of this thesis | 72 |
| It was found that the | 111 | The aim of this study | 55 |
| It is shown that the | 82 | It is argued that the | 38 |
| The aim of this research | 71 | This thesis is a study | 35 |
| The work presented in this | 69 | This thesis is concerned with | 35 |
| The aim of this thesis | 59 | The purpose of this thesis | 34 |
| The results show that the | 52 | The thesis is divided into | 30 |
| It has been shown that | 49 | The second part of the | 29 |
| This thesis describes the development | 47 | The aim of this research | 26 |
| It is found that the | 41 | Semi-structured interviews were conducted with | 25 |
| The second part of the | 39 | The purpose of this study | 24 |
| The work described in this | 38 | The thesis argues that the | 23 |
| This thesis focuses on the | 35 | This thesis focuses on the | 21 |
| The main objective of this | 32 | The findings are discussed in | 20 |
| In the second part of | 32 | This thesis sets out to | 20 |
| The work in this thesis | 32 | This study aimed to explore | 19 |
| The research presented in this | 32 | The main aim of this | 19 |
| The aim of this work | 31 | It is suggested that the | 19 |
| It is concluded that the | 30 | This is followed by a | 17 |
| The first part of the | 30 | The thesis concludes with a | 17 |
| It has been found that | 30 | The primary aim of this | 15 |
| It was also found that | 29 | The aim of the thesis | 15 |

The most frequent lexical bundle in the sub corpus of Physical Sciences corpus is *This thesis is concerned with,* which recurred 124 times per million words. The same bundle ranked fifth in the list of most frequent bundles in the Arts and Humanities corpus, with 35 recurrences per million words. This bundle is followed a *n* or *verb+ing,* forming a complete structure in the corpus, as shown follow:

This thesis is concerned with  (124)

This thesis is concerned with the application of a welding robot in a shipyard.

This thesis is concerned with ensuring high energy resolution from the swept charge device (SCD) CCD54, essentially a non-pixellated version of the charge coupled device (CCD), for use in the Chandrayaan-1 X-ray Spectrometer (C1XS).

This thesis is concerned with the measurement and interpretation of isotope shifts.

This thesis is concerned with the analysis of the main combustion products CO, CO2 and smoke from EVA nanocomposites, EVA and PVC cable materials under various ventilation conditions.

This thesis is concerned with investigating and understanding the role and consequence of different modelling options and parameter estimation options for modelling a complex plant.

This thesis is concerned with determining the possibility of the existence of subsonic free surface waves in piezoelectric half-spaces, and of subsonic free edge waves in elastic laminated semi-infinite plates, where the material constants are not presumed to possess any particular special form of symmetry.

This thesis is concerned with understanding the issues in generating cost estimates at the conceptual design stage for Embedded Systems Design and Development (ES D&D), based on specifications.

This thesis is concerned with the sheared flow of medium dense slurries, consisting rigid particles in a Newtonian fluid.

This thesis is concerned with approximation on compact homogeneous spaces.

This thesis is concerned with the theoretical behaviour and interactions of quantum systems.

This thesis is concerned with the design and realisation of components for a new state of the art 94GHz Electron Spin Resonance (ESR) spectrometer capable of operating in both pulsed and CW modes.

The most frequent lexical bundle in the sub corpus of **Arts and Humanities corpus** is *The aim of this thesis,* which recurred 72 times per million words. The same bundle ranked fifth in the list of most frequent bundles in the **Physical Sciences corpus**, with 59 recurrences per million words. This bundle is followed a *infinitive verb,* forming a complete structure in the corpus, as shown follow:

The aim of this thesis  (72)

The aim of this thesis is to cast new light on such questions by approaching them from within the cognitive framework of Sperber and Wilson's Relevance Theory.

The aim of this thesis is to give a systematic description of the acquisition of Persian and English syntax by a bilingual child.

The aim of this thesis is to understand some of the processes involved in the selection of appropriate tasks, and in the selection of correct responses to those tasks.

The aim of this thesis is to add to the growing body of knowledge about the effects of local and national events on the survival and fortunes of individual families and to explore the contribution of these families to the political scene.

The aim of this thesis was to explore the role of relationships, motivations, and meanings in dementia caregiving.

The aim of this thesis is to trace a 'language of looking' in some of Virginia Woolfs texts.

The aim of this thesis is to explore the activities of American women's associations in the international realm.

The aim of this thesis is to study in entirety the group of large colour prints which William Blake made between 1795 and 1805.

The aim of this thesis is to consider how Nawal El Saadawi's literary work can be examined on a theoretical level and to locate a possible theoretical framework, in her feminist texts, through which to explore the subjective identity of Arab women and the authenticity and creativity of their discourses.

The aim of this thesis is to investigate spatial and temporal aspects of agricultural change in southern Derbyshire between 1770 and 1870.

The verbal expressions used in the multiword units were also of great interest. The resulting bundles from the analysis of the sub corpus of **Physical Sciences corpus** included 11 main verbal expressions, namely *This thesis is concerned with, It was found that the It is shown that the, The results show that the, It has been shown that, This thesis describes the development, It is found that the, This thesis focuses on the, It is concluded that the,*

A Corpus-Based Comparative Study of Lexical Bundles in
PhD Dissertation Abstracts in Physical Sciences and Arts and
Humanities

*It has been found that, It was also found that.* Going through the Iranian corpus, one cannot fail to notice the greater number of verbal expressions within the identified bundles. The most commonly used verbal expressions by Iranian writers include: *It is argued that the,This thesis is concerned with,The thesis is divided into,Semi-structured interviews were conducted with,The thesis argues that the,This thesis focuses on the,The findings are discussed in,This thesis sets out to,This study aimed to explore,It is suggested that the,This is followed by a,The thesis concludes with a.*

The second round of analysis was to compare lexical bundles in the middle of the sentences . For example, *Overall the findings confirmed that in a simulated environment, propensity to violence in a UK sample was affected* <u>in the context of</u> *the variables tested. in the*

*context of* is a typical lexical bundle in the middle of the sentence.

The most frequent lexical bundle in the sub corpus of Physical Sciences corpus is *of this research is to,* which recurred 129 times per million words. The same bundle ranked 12th in the list of most frequent bundles in the Arts and Humanities corpus, with 39 recurrences per million words, sharply less frequently used compared to the Physical Sciences corpus. An apparent difference lies in the use of the word of "study" and "research". It turned out that the word "study" is preferred in Arts and Humanities Phd papers in such bundles as *of this study is to, of this study was to, : a study of the,* whereas "research" appears frequently in Physical Sciences Phd papers, as in bundles of *of this research is to, of this research was to.*

**Table 3**
**20 Most Frequent Bundles in the Middle of Sentences within the Two Corpora**

| Physical sciences corpus | Frequency | Arts and humanities | Frequency |
|---|---|---|---|
| of this thesis is to | 129 | of this thesis is to | 145 |
| of this research is to | 79 | of this study is to | 69 |
| it was found that the | 59 | in the context of the | 65 |
| in the form of a | 55 | the ways in which the | 61 |
| to investigate the effect of | 52 | are discussed in relation to | 55 |
| of this work is to | 50 | at the end of the | 51 |
| that can be used to | 49 | with particular reference to the | 49 |
| it is shown that the | 45 | of this thesis was to | 49 |
| that it is possible to | 45 | of this study was to | 48 |
| presented in this thesis is | 42 | the extent to which the | 47 |
| has been shown to be | 41 | the first half of the | 41 |
| as a result of the | 41 | of the ways in which | 41 |
| to the development of a | 41 | of this research is to | 39 |
| of this study was to | 40 | and the ways in which | 38 |
| the design and implementation of | 40 | in the light of the | 36 |
| of this research was to | 40 | : a study of the | 33 |
| as a function of the | 39 | within the context of the | 33 |
| one of the most important | 38 | aim of this thesis is | 31 |
| is one of the most | 38 | the way in which the | 30 |
| this thesis is concerned with | 38 | as a result of the | 29 |
| in order to determine the | 37 | discussed in relation to the | 27 |
| describes the development of a | 36 | one of the most important | 26 |

## 4.2 Structural Classification of Lexical Bundles

The identified lexical bundles were consequently classified into one of three main categories introduced by Biber et al. (1999). As is characteristic of academic prose and the research article register, in both sub-corpora, the majority of bundles were phrasal. From the list of bundles found in the Physical Sciences corpus, 14 were phrasal; on the other hand, 23 phrasal bundles were observed in the Arts and Humanities corpus of abstracts.

The second most common type of lexical bundle in corpus was the verb phrase fragments, among which 5 lexical bundles in the Arts and Humanities corpus and 7 in the Physical Sciences corpus, either with or without a preceding subject. Finally, verb phrase element followed by a dependent clause was only found in the beginning of the sentence, (8 and 3 instances respectively). It is worth mentioning that impersonal sentence such as "it is +past participle of verb+that" is mostly used in this category,

especially in the sub-corpus of Physical Sciences. Table 4 below shows the number and percentage of the different categories of bundles in the analyzed corpora.

It appears that for verb-containing bundles, past tense, such as " was", is frequently used in the Physical Sciences corpus, totaling 279 occurrences, while past tense appeared 137 times in the Arts and Humanities corpus, significantly less than the Physical Sciences corpus.

**Table 4**
**Types and Categories of Lexical Bundles in the Two Corpora**

| | Verb phrase fragments | | | Verb phrase elements +dependent clauses | | | Phrasal components | | |
|---|---|---|---|---|---|---|---|---|---|
| | Beginning | Middle | Total/percentage | Beginning | Middle | Total/percentage | Beginning | Middle | Total/percentage |
| Physical sciences | 2 | 13 | 15/40% | 8 | 0 | 8/21% | 10 | 4 | 14/39% |
| Arts and humanitie | 9 | 7 | 16/38% | 3 | 0 | 3/7% | 9 | 14 | 23/55% |

## 4.3 Functional Classification of Lexical Bundles

Bhatia (1994) suggests that research article abstracts provide readers with information about various aspects of the research article, including what the author did during the study; how it was done; what the author discovered; and what conclusions were drawn from the findings. Just like other sections of the research article, the abstract consists of a number of steps and moves. Bhatia (1994) found that the genre of research article was comprised of four moves: **Introducing the purpose, Describing the methodology, Summarizing the results, and Presenting the conclusions.**

A functional analysis of the lexical bundles shows that in both Physical Sciences and Arts and Humanities PhD theses abstracts, the largest number of bundles was used to introduce the purpose of the study. A number of examples of such bundles in context are provided below:

**This thesis is concerned with** the application of a welding robot in a shipyard.( Physical Sciences Corpus)

**The aim of this research** project is to explore a new technique .( Physical Sciences Corpus)

**The aim of this thesis** is to cast new light on such questions by approaching...(Arts and Humanities Corpus)

**The aim of this study** is to examine the ideology which informed Greek depictions of slaves. (Arts and Humanities Corpus)

The second most common function for the lexical bundles found was summarizing the results and presenting the conclusions. These two functions are very similar in nature and tend to overlap in terms of the linguistic features and lexical bundles they possess. For examples:

**It was found that the** best fitting equation was a sigmoid with 2 adjustable parameters that control the slope of the sigmoid and its offset point. (Arts and Humanities Corpus)

**The findings of this study** illustrate that participants constructed intuition as a valid concept with a significant role in clinical psychology practice. (Arts and Humanities Corpus)

Finally, the least number of lexical bundles in either corpus was devoted to describing the methodology of the study. Interestingly, it is found in this study that Physical Sciences PhD tend to use certain formulaic sequences more frequently and repetitively than Arts and Humanities PhD, because they view them as reliable 'safety nets' which can be confidently used, especially at times of uncertainty. In other words, Physical Sciences PhD is not as confident as Arts and Humanities PhD in terms of writing skills, so lexical bundles could be viewed as a form of compensation strategy for writing.

Another interesting finding is that most of the lexical bundles found in the Physical Sciences corpus were clearly used for achieving one of the abstract moves, while a considerable portion of bundles in the Arts and Humanities corpus (e.g. in the context of the, the ways in which the, are discussed in relation to, at the end of the, with particular reference to the, etc.) did not belong to any particular move and could generally be used for various purposes. This gives further plausibility to the previous claim that the use of lexical bundles in the writing may be a writing strategy, by means of which writers can achieve the move structure of any given genre more easily and with greater clarity. These lexical bundles function as 'genre frames', serving a dual purpose: First, they assist writers in more easily fulfilling genre expectations (e.g. introducing the purpose of the study, summarizing the results) and adding to the clarity of the passage, just as discourse markers enhance a reader's comprehension of a text. Second, they reduce the cognitive load imposed on writers by providing ready-made sequences stored as prefabricated units of language. However, Arts and Humanities PhD theses are more free and open in writing form and pattern.

The classification of lexical bundles based on Biber et al.'s (1999) taxonomy revealed that Physical Sciences PhD writers used considerably more verb phrase fragments and verb phrase elements followed by dependent clauses. Biber et al. (2011) claim that clausal subordination is more commonly observed in academic conversation compared to academic writing, which tends to include more noun phrase constituents and complex phrasal structures. According to these findings, Physical Sciences PhD writers, despite their high level of English language proficiency, tend to write in an academically conversational style, depending heavily on clausal

A Corpus-Based Comparative Study of Lexical Bundles in
PhD Dissertation Abstracts in Physical Sciences and Arts and
Humanities

elements and subordination. This finding is quite natural, since Physical Sciences PhD writers, as they may not be as confident in writing as Arts and Humanities PhD writers, attach more importance to the form of academic writing, with its reliance on extended noun phrases and limited use of verbs. This can only be seen in the academic writing register, which strives for greater brevity and density, conveying more information in fewer words.

## CONCLUSION

This study investigated the use of 5-word lexical bundles in PhD theses abstracts written by native English speakers in the field of Physical Sciences and Arts and Humanities. Through a corpus-driven approach to linguistic analysis, the two groups of abstracts were compared in terms of the number, type, and communicative function of lexical bundles with reference to the move-structure of abstracts. Results showed that Physical Sciences PhD writers tend to use more multi-word expressions in their writing, and also more commonly employ those expressions for serving specific genre expectations. Writing instructors can use the findings of this study to further emphasize the role of lexical bundles for their students, explaining to them the various ways these prefabricated language patterns can be used to achieve moves and steps, consequently aiding target readers in understanding their passage and heightening 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 more prominent use of clausal fragments and subordination in abstracts by Physical Sciences PhD authors is interesting. Other written registers of English, such as fiction and journalistic texts, are different in that they use more extensive subordinate clauses and rely less on noun phrases and nominalizations. The fact that abstracts by Physical Sciences PhD authors in this study did not highlight noun phrase structures as in the case of Arts and Humanities PhD abstracts could possibly mean that Physical Sciences PhD writers, despite their relatively high level of proficiency as published authors in the field of Physical Sciences, followed a general formal writing style compared to Arts and Humanities PhD writers.

The findings also have implications for writing instructors active 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.

The relationship between the lexical bundles found in this study and their functioning as genre-frames could also be of great instructional value. Since bundles such as *The aim of this study* are immediately associated with a particular move in abstracts, we could also consequently claim that readers encountering this bundle in an abstract would also more readily identify its aim. Therefore, it is clearly an advantage for any writer to make use of high-frequency bundles belonging to a specific genre move. Lexical bundles, like other multi-word expressions, are known to be stored holistically. This results in quicker accessibility and easier processing. Not only will this allow the writer to communicate the intended message more efficiently and with greater ease, but it would also make it easier for the reader to grasp the writer's intention. Future studies could possibly look into the frequency and function of lexical bundles used by learners/writers from other L1 backgrounds or other parts of the research article. The techniques through which lexical bundles can be taught or brought to learners' attention would also be an interesting topic for further investigation.

## 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.

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.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

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

Biber, D. (2004). Lexical bundles in academic speech and writing. In B. Lewandowska- Tomaszcyk (Ed.), *Practical*

*applications in language and computers—PALC 2003* (pp.165-178). Frankfurt is Main: Peter Lang.

Biber, D. (2006). *University language.* Amsterdam: 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., 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.

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

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.

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

Crossley, S., & Salsbury, T. L. (2011). The development of lexical bundle accuracy and production in English second language speakers. *International Review of Applied Linguistics in Language Teaching, 49*(1), 1-26.

Deane, P., & Quinlan, T. (2010). What automated analyses of corpora can tell us about students' writing skills. *Journal of Writing Research, 2*(2), 151-177.

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.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Language Acquisition, 24*(2), 143-188.

Ellis, N. C. (2008). Phraseology: The periphery and the heart of language. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp.1-13). Amsterdam: John Benjamins.

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.

Gries, S. T. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics, 13*(4), 403-437.