# A STUDY OF LEXICAL BUNDLES IN ACADEMIC LECTURES IN THE BASE CORPUS

KARIMA IBRAHIM EL TWATI

FACULTY OF LANGUAGES AND LINGUISTICS
UNIVERSITY OF MALAYA
KUALA LUMPUR

2019

# A STUDY OF LEXICAL BUNDLES IN ACADEMIC LECTURES IN THE BASE CORPUS

**KARIMA IBRAHIM EL TWATI**

**DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ENGLISH AS A SECOND LANGUAGE**

**FACULTY OF LANGUAGES AND LINGUISTICS**
**UNIVERSITY OF MALAYA**
**KUALA LUMPUR**

**2019**

**UNIVERSITY OF MALAYA**

**ORIGINAL LITERARY WORK DECLARATION**

Name of Candidate: KARIMA IBRAHIM EL TWATI

Matric No: TGB 150001

Name of Degree: MASTER OF ENGLISH AS A SECOND LANGUAGE

Title of Project Paper/Research Report/Dissertation/Thesis ("this work"):

**A STUDY OF LEXICAL BUNDLES IN ACADEMIC LECTURES IN THE BASE CORPUS**

Field of Study: Discourse Analysis

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

**A STUDY OF LEXICAL BUNDLES IN ACADEMIC LECTURES IN THE BASE CORPUS**

## ABSTRACT

Lexical bundles are combinations of more than two words which co-occur frequently in a given register (Biber, Johansson, Leech, Conrad, & Finegan, 1999). They lead to coherence in texts or speeches and play a key role in fluent linguistic production. Recent decades have witnessed an increasing body of research on lexical bundles, however there is still a question of whether these expressions are used differently in academic speech of different disciplinary divisions. To this aim, this study investigates and compares the frequencies, structures, and discourse functions of the most frequently occurring four-word lexical bundles in academic lectures across two broad disciplinary groups, namely Arts and Humanities and Life and Medical Sciences. The comparative study was run on a 594,878 corpus of academic lectures (40 from each academic division). The lectures were sourced from British Academic Spoken English (BASE) corpus. The most frequent four-word bundles were identified in each corpus using (AntConc). Then the structural and functional taxonomies proposed by Biber, Conrad, & Cortes (2004) were used as analytical frameworks to group lexical bundles in terms of their grammatical types and the discourse functions they serve. Primary findings revealed some variations between the two corpora in relation to the distributional patterns of the target bundle. In addition, the two groups of lecturers also showed different tendencies towards the selection of grammatical types to form lexical bundles and the functions that the bundles carried out in academic lectures. The results suggest that the selection of bundle types and the way they are used to fulfill disciplinary functions in the academic lectures are to a large extent disciplinary-bounded. Some bundles were also found to be specific to each corpus. Disciplinary lectures appeared to have their own specific ways of selecting lexical

bundles to convey disciplinary materials in a way to be as comprehensible as possible for the audiences. Based on the obtained results, it can be suggested that lexical bundles are considered as a pivotal means in distinguishing the academic speech of different fields of studies. This study has implication in ESL/ EFL academic settings. Findings of these study open more windows to how lexical bundles and their communicative functions are employed in academic disciplinary lectures. Students who study in these disciplines could also benefit from findings of this research by being familiarized with the structural and functional characteristics of lexical bundles.

**Key words:** Academic lectures, disciplinary division, discourse function, formulaic language, lexical bundles.

**KAJIAN GUGUSAN LEKSIKAL DALAM KULIAH AKADEMIK**

**DI KORPUS BASE**

## ABSTRAK

Gugusan leksikal adalah gabungan lebih dari dua perkataan yang kerap berlaku dalam sesuatu daftar yang diberikan (Biber, Johansson, Leech, Conrad, & Finegan, 1999). Mereka membawa koheren dalam teks atau ucapan dan memainkan peranan utama dalam pengeluaran linguistik yang fasih. Beberapa dekad kebelakangan ini telah menyaksikan peningkatan penyelidikan mengenai gugusan leksikal, namun masih terdapat persoalan sama ada ekspresi ini digunakan secara berbeza dalam ucapan akademik dari disiplin yang berbeza. Untuk tujuan ini, kajian ini menyiasat dan membandingkan frekuensi, struktur, dan fungsi wacana yang terdiri daripada ikatan leksikal perkataan empat ayat leksikal yang paling sering berlaku dalam kuliah akademik dua kumpulan disiplin yang luas iaitu Sastera dan Kemanusiaan, dan Sains Kehidupan dan Perubatan. Kajian perbandingan ini dijalankan menggunakan sebanyak 594,878 perkataan korpus kuliah akademik (40 dari setiap bahagian akademik). Kuliah-kuliah ini diperoleh daripada korpus *British Academic Spoken English (BASE)*. Gugusan leksikal empat perkataan yang paling kerap telah dikenalpasti dalam setiap corpus menggunakan AntConc. Taksonomi struktur dan fungsian yang dicadangkan oleh Biber, Conrad, & Cortes (2004) telah digunakan sebagai kerangka analisis untuk mengumpul gugusan leksikal dari segi jenis tatabahasa dan fungsi wacana yang mereka layani. Penemuan utama mendedahkan beberapa variasi antara dua corpora berkaitan dengan corak distribusi gugusan sasaran. Di samping itu, kedua-dua kumpulan pensyarah juga menunjukkan kecenderungan yang berlainan terhadap pemilihan jenis tatabahasa untuk membentuk gugusan leksikal dan fungsi-fungsi yang dijalankan dalam kuliah akademik. Keputusan mencadangkan bahawa pemilihan jenis gugusan dan cara mereka digunakan untuk memenuhi fungsi disiplin dalam kuliah akademik adalah sebahagian besarnya dibatasi disiplin mengikut disiplin

tersebut. Beberapa gugusan juga didapati khusus untuk setiap korpus. Syarahan kelihatan mempunyai cara tersendiri mereka untuk memilih gugusan leksikal bagi menyampaikan bahan-bahan disiplin dalam cara yang dapat difahami oleh penonton. Berdasarkan hasil yang diperoleh, dapat dicadangkan bahawa gugusan leksikal dianggap sebagai alat penting dalam membezakan ucapan akademik dari berbagai pelbagai pengajian. Kajian ini mempunyai implikasi dalam penetapan akademik ESL / EFL. Penemuan-penemuan kajian ini membuka lebih banyak tingkap untuk bagaimana bungkusan leksikal dan fungsi komunikatif mereka digunakan dalam kuliah akademik. Pelajar yang belajar dalam bidang ini juga boleh mendapat manfaat daripada penemuan kajian ini dengan membiasakan diri mereka dengan ciri-ciri struktur dan fungsi gugusan leksikal.

Kata kunci: Ceramah akademik, pembahagian tatatertib, fungsi wacana, bahasa formulaik, kumpulan leksikal.

## ACKNOWLEDGEMENTS

Alhamdulillah for his blessings in giving me strength and health in completing this thesis. I would like to express my gratitude to the Ministry of Higher Education in Libya for giving me the opportunity to complete my Master's Degree in English as a Second Language at the Faculty of Languages and Linguistics at the University of Malaya. This has been a rewarding experience for me and my thanks go to all my professors at the Faculty of Languages and Linguistics at the University of Malaya for their highly valuable teachings and advice.

I am forever grateful to Dr. Sheena Kaur, my supervisor, without her this work would not have been possible. I have learnt a lot throughout the writing process of this thesis and I could never thank her enough for her time, patience, and guidance. I feel fortunate to have had the opportunity to work with her. I would also like to thank my thesis committee members, Dr. Siti Zaidah Zainuddin for her valuable advice and feedback. I am also thankful to Dr. Chau Meng Huat for his insightful advice, comments and suggestions.

Last but not least, special thanks to my beloved family for their constant prayers and encouragement. My heartfelt thanks go also to my husband Abdulmutaleb for his everlasting understanding, patience, and support.

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

## LIST OF SYMBOLS AND ABBREVIATIONS

| | | |
|---|---|---|
| BASE | : | British Academic Spoken English |
| ESL | : | English as a Second Language |
| EFL | : | English as a Foreign Language |
| EAP | : | English Academic Purposes |
| MICASE | : | Michigan Corpus of Academic Spoken English |
| AH | : | Arts and Humanities |
| LS | : | Life and Medical Sciences |
| LGSWE | : | Longman Grammar of Spoken and Written English Corpus |
| ESP | : | English for Specific Purposes |
| T2K_SWAL | : | TOEFL 2000 Spoken and Written Academic Language Corpus |
| COCA | : | Corpus of Contemporary American English |
| AWL | : | Academic Word List |
| BNC | : | British National Corpus |
| BAWE | : | British Academic Written English |
| MUET | : | Malaysian University English Test |
| TOFEL iBT | : | TOFEL Internet-based Test |
| L2 | : | Second Language Learner |
| ELF | : | English as a Lingua Franca |
| AELF | : | Academic Lingua Franca English |
| VP | : | Verb Phrase |
| NP | : | Noun Phrase |
| PP | : | Prepositional Phrase |

## LIST OF APPENDICES

# CHAPTER 1: INTRODUCTION

## 1.1 Background of the Study

Simply defined as extended collocations, lexical bundles were first found in The Longman Grammar of Spoken and Written English (Biber, Johansson, Leech, Conrad & Finegan, 1999). Biber et al., (1999) defined lexical bundles as "sequences of word forms that commonly go together in natural discourse" (p. 990). Lexical bundles were studied under different terms such as prefabricated routines and patterns (prefabs) (Erman & Warren, 2000); or chunks (Hyland, 2012). According to Biber et al. (1999), lexical bundles are word sequences that frequently happen to occur together unpredictably and do not follow any special sequence pattern *(*e.g., *do you want me* and *in the case of)*. However, lexical bundles have a crucial role in understanding and comprehending the meaning of particular contexts and greatly contribute to the coherence of the text (Hyland, 2008).

The important role of lexical bundles in written or spoken discourse was broadly investigated in many studies (Biber et al, 1999; Biber, Conrad & Cortes, 2004; Cortes, 2004, 2006; Biber & Barbieri, 2007; Hyland, 2008a, 2008b). These studies showed the significance of lexical bundles as a primary source of coherence in academic texts of various discourse genres/registers. Many of the investigations carried out on lexical bundles concentrated on functional and structural analysis of these recurring expressions (Cortes, 2004; Chen & Baker, 2010; Hyland, 2008; Strunkyt & Jurkūnait, 2008; Bal, 2010; Adel & Erman, 2012). Structural classifications of bundles were first mentioned by Biber and his colleagues (1999). They discovered that lexical bundles are structurally complex, generally incomplete and unfixed and they categorized them into broad structural classifications (e.g., noun phrase + of phrase fragments, other noun phrase expressions, prepositional phrase +of, prepositional phrase expressions, passive + prep phrase fragment, anticipatory it + verb/ adj, be + noun/adjectival phrase and others) (Biber

et al., 1999). With the identification of the grammatical forms, there was a need to have other frameworks for analyzing the functions of lexical bundles within a text.

### 1.1.1 The Concept of Lexical Bundles

With the advancement in computer technology, the analysis and calculation of multi-word units have become much simpler (Jablonkai, 2012). The Longman Grammar of Spoken and Written English (Biber et al., 1999) was the first grammar published within a framework which focused on lexical bundles. A number of scholars provided operational definitions of lexical bundles in their related research (Biber et al., 2004; Cortes, 2004; Hyland, 2008). A series of comprehensive studies conducted by Biber et al., (2004) on the use of lexical bundles in a variety of written and spoken registers were considered pioneering studies for the understanding of lexical bundles and their grammatical structures and discourse functions.

In their first work, Biber et al. (1999) explored lexical bundles in conversations and academic prose where they defined lexical bundles as "recurrent expressions, regardless of their idiomaticity, and regardless of their structural status" (p. 990). They are word combinations that co-occur frequently and are basically identified empirically rather than intuitively. The research was expanded later by Biber and Conrad (1999) and Biber et al. (2003, 2004), comparing the use of lexical bundles across a variety of other registers such as conversations, textbooks and classroom teaching. Biber and Conrad (1999, p. 183) reworded their earlier definition and defined lexical bundles as "multi-word expressions which occur frequently and with accidental sequences of three or more word such as: *in the case of* and *on the other hand*. Lexical bundles are frequently-occurring chuncks of words which do not have any special sequence, that is, in most cases; words come one after another by chance. These expressions are incomplete structural elements, but serve as building blocks of language (Biber et al., 1999; Biber & Conrad, 1999). Hyland (2008) also states that lexical bundles have an important role in creating coherence "extended

collocations which appear more frequently than expected by chance, helping to shape meanings contributing to our sense of coherence in a text" (p. 41). Lexical bundles have three essential characteristics that differentiate them from other multi-word combinations which include frequency of occurrence, fixedness, and incompleteness.

### 1.1.2 Characteristics of Lexical Bundles

The first characteristics regarded frequency as a determining criteria for studying lexical bundles, as it is a "reflection of the extent to which a sequence of words is stored and used as prefabricated chunks" (Biber et al., 2004, p.376). However, frequency is only one condition under which a sequence becomes prefabricated, for example, sequences containing idiomatic meanings are rarely prefabricated.

The second characteristic is that lexical bundles are not structurally complete. They mostly link two units, for example, *there are a lot* and *of the most important.* Biber et al. (1999) discovered that in conversations, only 15% of lexical bundles are structurally complete and 5% of those found in academic prose can be regarded as structurally complete. In other words, the majority of lexical bundles are themselves incomplete and therefore usually bridge two separate structural units, beginning at a phrase or clause boundary, with the final words of the bundle representing the first element of a second structural component (Biber, 2006). Despite being structurally incomplete and non-idiomatic, lexical bundles play a vital role in discourse: they serve as "a kind of pragmatic 'head' for larger phrases and clauses, where they function as discourse frames for the expression of new information" (Biber & Barbieri, 2007, p. 270).

The third characteristic is idiomaticity, lexical bundles are not idiomatic in their meaning. They are "semantically transparent and formally regular" (Hyland, 2008a, p. 6). For example, the general meanings of lexical bundles like *these results suggest that* and *as can be seen* are fully extracted from the single words that compose them. Further, lexical bundles are generally not "perceptually salient" (Biber & Barbieri, 2007), because they are identified primarily based on their frequency rather than their structure. That is, a lexical bundle functions as an entire unit, unlike idioms, its meaning can be well understood from the words that make up the bundle. Their meaning is derived from the words they are composed of. These would lead to a notable consideration in view of this fact that lexical bundles appear to bridge syntactic boundaries and are generally not idiomatic in meaning. Bundles such as *when you talk about* and *at the same time* are transparent in meaning from their component parts (Biber, 2006), that is "their meaning is transparent" (Cortes, 2004, p.400). They can be understood easily based on the translation of the individual words that they are made of.

In order to identify lexical bundles in text or speech, two different criteria are required. The first is the frequency-cut off. Biber et al. (1999) mentioned that for a word combination to qualify as a lexical bundle, it must appear at least 10 times per 1 million words and spread across a minimum of 5 texts. The most common bundles in a corpus can have a much higher frequency that reach over 100 times per million words (Cortes, 2004). On the other hand, even though there is a consensus on the frequency as the primary criterion for recognizing lexical bundles, the cut-off applied in lexical bundles varies from study to study. For instance, Biber (2006), Cortes (2004) or Hyland (2008a, 2008b) used a greater frequency than Biber et al. (1999), setting their cut-off point at 20 times in every 1 million words. Other different studies (e.g., Biber et al., 2004; Biber & Barbieri, 2007) applied a higher frequency cut-off point of 40 times per every 1 million

words. The concept is that the greater the frequency, the more representative a particular register the identified bundles are.

Another parameter is dispersion or distributional patterns that must be satisfied for a recurring lexical sequence to qualify as lexical bundles. Dispersion means that it must occur in multiple texts within a register as well as according to Hyland (2012), refers to their use by various language users. Biber et al.'s (1999) state that lexical bundles must be spread across at least five different texts in each register. This criterion is important because it avoids focusing on idiosyncratic uses in the corpus under consideration. Also, the minimum dispersion can vary across studies. For example, that patterns must occur in a particular number of texts in the corpus like three to five texts (e.g., Biber & Barbieri, 2007) or about 10% of texts (Hyland, 2008a) in order to prevent quirks of individual writers or speakers, but for small spoken corpora, a raw cut-off frequency point often ranges from 2 to 10 (e.g., Altenberg, 1998; De Cock, 1998).

Finally, researchers must decide on the length of bundles they select (Biber et al., 2004; Cortes, 2004; Hyland, 2008; Dontcheva-Navratilova, 2012; Grabowski, 2013). Three-word lexical bundles are very common, and seem not to be very interesting, while five and six-lexical bundles are somewhat infrequent and often incorporate shorter ones. Four-word lexical bundles tend to be most often explored, perhaps because they are about 10 times more frequent than five-word sequences and have a broader variety of functions and structures to analyze. Biber, et al. (1999) in fact, suggest that four-word bundles and over "are more phrasal in nature and correspondingly less common" (p. 992). According to Hyland (2008), "four-word lexical bundles are far more common than five-word strings and offer a clearer range of structures and functions than three-word bundles" (p. 8).

### 1.1.3 The Importance of Lexical Bundles

Lexical bundles have attracted researchers' attention due to their functional role to the coherence of various written and spoken texts. Lexical bundles are primary building units of discourse (Biber & Barbieri, 2007), and aid writers to recognise within a particular discourse community (Adel & Erman, 2012). Lexical bundles can be stored as "unanalyzed multi-word chunks, rather than as productive grammatical constructions" (Biber et al., 2004, p. 400) and recognized as "wholes and not as strings of individual words" (Perez-Llantada, 2014, p. 83), thus they can be automatically extracted making writing and speech more fluent. Additionally, by signalling the suitable use of the disciplinary resources, lexical bundles enable writers to show solidarity with colleagues (Cortes, 2006) and to build a disciplinary competent voice (Hyland, 2008a; Pang, 2010), that help learners to comprehend academic textbook language (Wood & Appel, 2014).

Hyland (2012) claims that lexical bundles are crucial to speakers and writers for three main reasons: "(1) their repetition offers users (and particularly students) ready-made sets of words to work with, (2) they help define fluent use and therefore expertise and legitimate disciplinary membership, [and] (3) they reveal the lexico-grammatical community-authorized ways of making-meanings" (p.153). These advantages of lexical bundles have motivated researchers to analyze lexical bundles across different disciplines, registers, and genres. Therefore, researchers emphasize the importance of studying these bundles as they are important building blocks which increase discourse coherence (Hyland, 2008). Finally, lexical bundles can distinguish between expert use and novice use of spoken and written language in various contexts, aiding members of a discourse group to show harmony with all other members (Wray, 2006; Huang, 2015).

## 1.2 The Study of Lexical Bundles in Academic Discourse

Academic genres have attracted a considerable amount of interest in English for Academic Purposes (EAP). Past researchers (Cortes, 2002, 2004; Biber et al., 2004; Nesi & Basturkmen, 2006; Hyland, 2008) have demonstrated that in-depth study into lexical bundles revealed genre-specific features in language use (Greaves & Warren, 2010). Biber and Barbieri (2007, p. 265) pointed out that "each register employs a distinct set of lexical bundles, associated with the typical communicative purpose of the register". Thus, lexical bundles have been examined in a broad range of specialized registers. In academic written discourse, various studies were conducted to explore lexical bundles used in written academic registers such as in textbooks and research papers (Biber et al., 2004; Cortes, 2004; De Cock, 2004; Biber & Barbieri, 2007; Hyland, 2008; Adel & Erman, 2012).

Most studies on registers investigated lexical bundles with different purposes and various registers. Some of these studies explored and compared lexical bundles in two registers, i.e., academic writing and speech (e.g., Biber et al, 1999; Biber et al., 2004; Biber & Barbieri, 2007). Others looked at academic vs. non-academic registers (e.g., Biber & Barbieri, 2007), and professional writers vs. students (e.g., Scott & Tribble, 2006; Hyland, 2008a; Erman, 2009; Adel & Erman, 2012), L1 vs. L2 writing (e.g., De Cock, 2003; Paquot, 2008; Chen & Baker 2010; Pe´rez-Llantada, 2014), as well as academic disciplines (e.g., Cortes, 2002, 2004; Hyland, 2008b). Some studies focused on one register only, either academic writing (e.g., Wary & Berkins, 2000; Simpson-Vlach & Ellis, 2010; Liu, 2012) or academic spoken discourse (e.g., Nesi & Basturkmen, 2006; Csomay & Cortes, 2007, 2010; Neely & Cortes, 2009; Herbel-Eisenmann & Wagner, 2010). There are other studies that explored lexical bundles in languages other than English or comparing two languages (e.g., English vs. Spanish). Investigating lexical

bundles for pedagogical reasons also can be carried out. The results of these investigations have provided new contributions in the field of English for Academic Purposes (EAP).

However, research on lexical bundles in academic spoken genres has received less attention compared to academic written genres (Biber et al., 2004; Biber, 2006; Biber & Barbieri, 2007; Cortes & Csomay, 2007). Studies by Biber et al., (2004) and Biber & Barbieri (2007) have studied a number of genres/ registers including instructional contexts as in study groups and classroom teaching, and non-instructional contexts as in student advising, class management and university service encounters. Lexical bundles were also investigated in interviews (De Cock, 2004; Larsson, 2011). Furthermore, a few related studies on spoken academic genres involved the study of academic lectures (Nesi & Basturkmen, 2006; Neely & Cortes, 2009; Deroey & Taverniers, 2012; Csomay, 2013; Kashiha & Chan, 2013, 2014; Kashiha, 2015).

### 1.2.1 The Study of Lexical Bundles in Academic Lectures

As opposed to written discourse, lexical bundles or other multi-word expressions in spoken discourse had less focus from scholars. The studies that analyzed these clusters seemed to be growing as a result of the advances in technology and the emergence of new methodologies and led to the accessibility to large online spoken corpora (Nesi & Basturkmen, 2006; Neely & Cortes, 2009; Csomay, 2013; Kashiha & Chan, 2013, 2014; Kashiha, 2015).

Biber et al. (2004) argue that lexical bundles function as speech organizational structures that help audience anticipate and understand information that is delivered to them. Also, Nesi and Basturkmen (2006) explore the role that lexical bundles have in constructing cohesion in a total of 160 different academic lectures and reveal that lexical bundles have a crucial role in achieving cohesion. In academic lectures, a variety of lexical bundles are used to interpret information that cater specifically to a discipline. It

is therefore sensible to claim that lexical bundles behave differently across disciplines and contexts (Hyland, 2002; 2006). Hyland and Tse (2007, p. 240) state that "all disciplines shape words for their own uses" and thus defend the discipline-specific approach to EAP.

However, there is increasing evidence that discipline matters in academic speech (e.g., Ädel, 2008; Schleef, 2008; Kashiha & Chan, 2014; Wang, 2017). Kashiha & Chan (2014) and Wang (2017) identified important differences in the frequency and structural forms of lexical bundles in many different academic disciplines. Despite the significant emphasis on the usage of the bundles in EAP, a crucial question still remains about the degree to which they vary in the use of academic lectures in terms of discipline-specific discourse functions.

### 1.2.2 Lexical Bundles and Discipline Variation

Research works investigated the similarities and differences of lexical bundles in various genres and within a single discipline or across different disciplines in academic written discourse (Cortes, 2004; Hyland, 2008a, 2008b; Strunkyt & Jurkūnait, 2008; Jalali, 2009; Chen & Baker, 2010; Valipoor, 2010; Parvizi, 2011; Rafiee, Tavakoli & Amirian, 2011; Adel & Erman, 2012; Karabacak & Qin, 2013).

Variation between disciplines regarding lexical bundles has been of particular interest. However, the notion of lexical bundles in academic lectures have been explored from cross-disciplinary perspective and in relation to other linguistic features (Nesi & Basturkmen, 2006; Neely & Cortes, 2009; Csomay, 2013; Hernández, 2013; Kashiha & Chan, 2013, 2014a, 2014b; Kashiha, 2015). For instance, Nesi and Basturkmen (2006) examined 160 lectures from BASE and MICASE online corpora. Neely and Cortes (2009) focused their attention on MICASE online corpus by examining lexical bundles in the lecturers' speech and students to identify the functions of five lexical bundles. Csomay

(2013) studied the distributional patterns of the lexical bundle functions reported in previous studies (Biber et al., 2004) and explored how they are related to discourse structure. Studies by Kashiha and Chan (2013, 2014a, 2014b) explored lexical bundles in terms of frequency, structures and functions in academic lectures of BASE corpus in a variety of disciplines. Kashiha (2015) subsequently investigated and compared the frequency, functions, and structures of the most common four-word lexical bundles in academic lectures across three broad disciplinary groups taken from BASE corpus. It is therefore crucial that EAP researchers develop a reliable understanding of the nature of disciplinary variation.

Therefore, there is a need to carry out a corpus-based research to investigate the use of lexical bundles to provide a clearer picture of discourse use within different communities. This study aims to identify the frequently occurring lexical bundles in different academic lectures of two broad disciplinary divisions: Arts and Humanities, and Life and Medical Sciences to make a comparison of the functions of the different bundles which are seen in each discipline.

## 1.3 Problem Statement

Developing fluency and coherence in academic discourse, or more specifically, university lectures, is largely affected by the use of lexical bundles (Biber et al., 2004; Nesi & Basturkmen, 2006; Biber & Barbieri, 2007; Csomay, 2013; Kashiha & Chan, 2013; 2014; Kashiha, 2015). Previous studies focused on the varying uses of lexical bundles in relation to other linguistic aspects in academic lectures, Nesi and Basturkmen (2006) explored the cohesive role of the lexical bundles in 160 lectures; as well as another study by Neely and Cortes (2009), who examined the functional types of bundles in lecturer and student speech. Similarly, Csomay (2013) explored the correlation between the function of bundles and their position in discourse structures. Another study analyzed

lexical bundles quantitatively, with the results reporting only the frequency of the identified bundles without giving a comprehensive explanation on the reason behind the bundle use (Rafiee et al., 2011).

Significant research work has been conducted on the use of lexical bundles in academic lecture, yet studies on the use of lexical bundles in academic lectures in their broad disciplinary divisions are scarce (Kashiha & Chan, 2013, 2014a, 2014b; Kashiha, 2015; Wang, 2017, 2018). Therefore, it is noteworthy to point out the variations of the frequent lexical bundles in disciplinary domains that include a greater number of disciplines in larger corpora to show the extent of its use in terms of structure, type and function. This could help learners to gain communicative competence in their field of study, since each discipline employs a distinctive group of lexical bundles linked with the subject-matter of that particular discipline, and the typical communicative purposes of that genre (Wang, 2017).

The present research is motivated by few studies that include specific empirical research to analyze and compare the use of 4-word lexical bundles in academic lectures of different broad disciplinary divisions. There is a need to explore the lexical bundles within the academic lectures to find out how lectures organize their discourses differently in the different disciplines.

## 1.4 Research Objectives

The aim of this study is to explore the lexical bundles found in academic lectures. This research work aims to investigate the use of frequently occurring 4-word lexical bundles within academic lectures of two broad disciplinary divisions: Arts and Humanities, and Life and Medical Sciences so as to come up with empirical data on the possible similarities and variations regarding the frequencies, structural characteristics and the

discourse functions associated with the identified bundles. The two disciplinary divisions were selected on the basis of fulfilling the purpose to explore the use of lexical bundles in the two broad academic disciplinary divisions: Arts and Humanities (non-scientific disciplines) and Life and Medical Sciences (scientific disciplines).

## 1.5 Research Questions

The current study aims to answer the following questions:

1. What are the most frequent four-word lexical bundles found in the academic lectures of Arts and Humanities (AH), and Life and Medical Sciences (LS) disciplines?

2. How are the academic lectures in Arts and Humanities, and Life and Medical Sciences disciplines different or similar in terms of the structural characteristics and discourse functions of the lexical bundles used?

## 1.6 Significance of the Study

Research in the area of formulaic language and particularly lexical bundles are of great value in applied linguistics. By conducting comparative studies, scholars could better understand the types of language that different communities of users encounter, as lexical bundles are important indicators in determining whether language users are successful in a specific discourse community. It is expected that the results acquired from this study will provide scholars with a clearer picture of community-specific practices that will be obtained by studying the variation of lexical bundles in university lectures across broad disciplinary divisions.

More importantly, the results hold pedagogic implications in the field of EAP. This study will shed light into the possible variations of lexical bundles used in a variety of cross-disciplinary fields of studies, and will provide an effective pedagogy to EAP

teachers. Moreover, the list of formulaic sequences taken from this study could facilitate the incorporation of multiword expressions into teaching materials. In addition, the results of investigations are important to applied linguists, EAP teachers and materials writers who can employ these results to enhance their teaching and overall materials development.

To this end, this study aims at providing perspective on lexical bundles used in academic lectures in two broad disciplinary groupings, namely Arts and Humanities, and Life and Medical Sciences. The study has crucial implications for language use in particular for university lectures from different disciplines through specific textual features associated with lexical bundles realized in the content of discourse organization. In addition, learning the way lectures from different disciplinary backgrounds construct lexical bundles structurally and use them to convey specific disciplinary information and these communicative purposes related to their field of study is of great value for lecturers.

## 1.7 Limitations of the Study

The first limitation of this study is the size of the corpus. This study only analyzed 40 academic lectures from each disciplinary division which may not be adequate for generalizing the reported findings. Further studies with larger corpora could be conducted to have a more insightful picture of the usage of lexical bundles, their structures and discourse functions within academic disciplinary lectures.

The limitations also point to the need to further work on an analytical framework used for lexical bundles. Thus, further research could be motivated towards this area to have a more detailed analytical framework to provide a solid foundation in the analysis of lexical bundles.

## 1.8 Summary of the Chapter

This chapter started by presenting some background information regarding lexical bundles and their importance in academic university lectures. Next, the problem statement, research questions, objectives of the study, and significance of the study were presented. This is followed by providing information concerning the limitations of this research.

## 1.9 Organization of the Thesis

There are five chapters in this study. The first chapter introduced the background information, the scope, objectives, research problem, research questions and the significance of the research. Chapter Two reviews the relevant research conducted over the past two years. It also summarizes the literature on lexical bundles by placing the previous studies into categories according to their research goals. A critical review was conducted to highlight the research gap in past studies and present the possibility for further research. Chapter Three delineates the methodology applied in the present study. It offers detailed information on data retrieving, data processing, and data analysis.

Chapter Four reports a detailed description of the main results of the current study. It interprets the main findings with an in-depth discussion. It restates the research questions addressed in Chapter One and attempts to offer a thorough analysis on what has been discovered. The last chapter, Chapter five, ends the thesis with a brief summary of the current research, highlighting its merits as well as its limitations. Research and pedagogical implications, along with suggestions for future research, are stated at the end of the chapter.

# CHAPTER 2: LITERATURE REVIEW

## 2.1 Introduction

This chapter comprises the theoretical background and a discussion on the previous studies relating to lexical bundles in academic lectures. This chapter highlights the importance of lexical bundles used in academic lectures in different disciplines.

## 2.2 Theoretical Background

Researchers have shown a special interest in investigating multi word combinations that tend to co-occur for quite a long time. Researchers have used different terms to refer to multi-word combinations. Firth (1957) was one of the first linguists that drew attention to multi-word combinations as well as Jespersen (1924) and Firth (1957) called them as collocations. In the traditional or Chomskyan way of viewing language, a distinction is made between the lexis and grammar (Chomsky, 1957, 1965). Language texts were regarded as a series of open slots that allow creative uses and any word should have a chance to occur in each slot and the only restriction is the grammar (Sinclair, 1991). The idiom principle was coined by Sinclair (1991) to acknowledge what has been demonstrated in corpus studies (e.g., Biber et al., 1999; Erman & Warren, 2000; Meunier & Granger, 2008), namely that "a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices" (Sinclair, 1991, p. 110). According to Sinclair (1987, 1991), it is lexis, not syntax that is responsible for the organization and patterning of language whereas the role of syntax is diminished to that of a structure to which chunks are slotted. Sinclair's (1987) "idiom principle" claims that writers and speakers choose pre-constructed phrases to express units of meaning and they do not select the words that they use one at a time.

Research into lexical bundles followed the pioneering work of Altenberg (1993, 1998), who created a new methodology to identify frequency-defined recurrent word combinations and who combined structural and functional analysis in categorizing these multi-word combinations. Many researchers (e.g., Altenberg, 1998; Sinclair, 1991) and syllabus designers (Willis, 1990; Lewis, 1997) followed corpus-based approaches and pedagogically focused on these clusters. These clusters or collocations are "the relationship that a lexical item has with items that appear with greater than random probability in its textual context" (Hoey, 2005, p. 3). According to Hyland (2008) most clusters are a combination of two or more items that seem to co-occurrence together for a purpose and appear repeatedly across various texts. He defined these clusters as "words which follow each other more frequently than expected by chance" (p.5). Altenberg (1998) suggests that nearly 80 per cent of natural language could be patterned in this way. Such patterning leads researchers such as Sinclair (1991) and Hoey (2005) to replace traditional theories of grammar with new theories of language. That is to say, lexical choices are controlled by the slots which grammar makes available for them and systematically these lexis are structured through the use of repeated patterns. As Sinclair (1991, p. 108) states:

> By far the majority of text is made of the occurrence of common words in common patterns, or in slight variants of those common patterns. Most everyday words do not have an independent meaning, or meanings, but are components of a rich repertoire of multi-word patterns that make up a text. This is totally obscured buy the procedures of conventional grammar.

Throughout the history of language studies, there have been many investigations that focused on features of academic writing. As cited by Biber (2006), most of these studies focused on different aspects of academic written discourse such as stance expressions (e.g., Holmes, 1986; Myers, 1989, 1990; Hyland, 1994, 1998; Salager-Meyer, 1994; Crompton, 1997; Grabe & Kaplan, 1997; Charles, 2003; Silver, 2003; Varttala, 2003);

verb classes (e.g., Hunston, 1995); academic registers (e.g., Hewings, 2001; Flowerdew, 2002) and the discourse organization (e.g., Ferguson, 2001), to mention only a few. Academic vocabulary is one of the features that also attracted attention, and analyzing academic vocabulary was the purpose of numerous studies such as (Nation, 1990, 2001; Schmitt & McCarthy, 1997; Coxhead, 2000).

The study of multi-word combinations such as lexical bundles experienced dramatic growth recently (O'Keeffe, McCarthy & Carter, 2007). Lately, there was a shift from studying single lexical items to studying multi-word expressions. Therefore, studies moved their focus from formulaic expressions and begun to go beyond the analysis of single lexical items (Pawley & Syder, 1983; Nattinger & DeCarrico, 1992; Altenberg, 1998; Biber et al., 1999). Decarrico & Nattinger (1988) were among the first to investigate multi-word units in university lectures. But the units in the study were identified on the basis of the researchers' intuition.

Earlier studies emphasized the importance of these formulaic expression, in particular, lexical bundles in different registers, languages and contexts (Biber et al., 1999; Biber & Conrad, 1999; Biber et al., 2004). An important step in the investigation of lexical bundles was made by Biber and his colleagues (1999), where they defined the term lexical bundles as sequences of three or more words found in natural language either written or spoken. These bundles were identified empirically across registers and showed the statistical tendency of their frequency of occurrence.

Biber et al. (1999) and Biber and Conrad (1999) based their research studies on a large corpus of both British and Academic English in academic prose and conversation. Biber et al. (1999) and Biber and Conrad (1999) reported their major findings on 3-word to 5-word lexical bundles from a large corpus of 4.5 million words of American English conversation and another corpus of 5.3 million words of research articles and academic

books in the Longman Grammar of Spoken and Written English Corpus (LSWE Corpus). The methodology applied in these two studies, i.e., their definition of lexical bundles and the structural taxonomy, had been widely adopted by a number of subsequent studies. According to Biber et al. (1999), for items to be considered as lexical bundles, they must occur at least 10 times per 1 million words in a register and should be distributed in at least 5 different texts in order to avoid individual idiosyncratic use. To limit the scope of description, the researchers focused on two registers only: conversation and academic prose. Some striking differences in grammatical structures of lexical bundles were discovered after comparing the most common lexical bundles between these two registers.

Biber et al. (1999) stated that "both conversation and academic prose use a large stock of different lexical bundles" (p.993). First, conversation contained a larger amount of lexical bundles than those in academic prose. Just 15% of the lexical bundles in conversations were complete grammatical clauses or phrases. In academic writing, less than 5% of the lexical bundles were complete structural units. Second, in conversation, 90% of the lexical bundles were parts of interrogative or declarative clauses and the majority of them include part of a verb phrase, that is, VP-based units and dependent clause bundles. About 50% of the four-word lexical bundles began with a personal pronoun + verb phrase *(*e.g., *I thought that was* and *I don't know why).* In contrast, over 60% of the lexical bundles in academic prose were nominal chunks, i.e., parts of noun phrases and prepositional phrases. Take four-word lexical bundles for example; there were no common VP-based bundles found in the corpus of academic writing, and only a few common bundles incorporated dependent clauses *(*e.g., *the fact that the, it is possible to* and *the way in which).* Third, lexical bundles in conversation usually bridged two clauses *(*e.g., *well that's what you* and *you want to say)*, while most of the bundles in academic prose linked between two phrases *(*e.g., *the result of the* and *in the case of).*

About 70 per cent of the common bundles in academic writing consisted of noun phrase expressions (e.g., *the nature of the*) or a sequence that bridged two prepositional phrases *(*e.g., *as a result of).*

A functional taxonomy of lexical bundles was formulated by Biber et al., (2004) and Biber and Barbieri (2007). Biber et al. (2004) in another extensive study compared lexical bundles in textbooks and classroom teaching with those in conversation and academic writing. They presented a functional taxonomy, identifying three primary discourse functions for the target lexical bundles: 1) stance expressions *(*e.g., *I don't want to* and *I don't know what);* 2) discourse organizers *(*e.g., *you know what I* and *what I want to);* and 3) referential expressions *(*e.g., *at the end of* and *the nature of the).* Compared to conversations, more bundles were used for stance and discourse organizing functions in classroom teaching; whereas, in academic prose, most bundles were used for referential functions. The number of lexical bundles doubled in university lectures in comparison with conversation. University lectures had four times the number of lexical bundles as were found in textbooks.

In addition, using the T2K-SWAL, Biber and Barbieri (2007) investigated the use of lexical bundles in instructional registers and non-instructional registers. More broadly, Biber and Barbieri (2007) looked at lexical bundles in a wider range of written and spoken university registers. In their study, they stated that lexical bundles were common in all university registers. They were specifically widely employed in spoken registers, while bundles rarely occurred in academic written registers. The non-instructional /non-academic written registers (particularly written course management) used more lexical bundles than any other university registers. Regarding the discourse functions, lexical bundles that expressed stance were most commonly found in spoken registers, while lexical bundles that conveyed referential expressions were more common in written

university registers. Biber and Barbieri argued that one of their main findings suggested that both mode and communicative purpose had an effect on the use of lexical bundles. Unlike previous studies which showed that lexical bundles were more common in spoken registers than in writing, while lexical bundles were very common in instructional written registers such as course syllabuses.

However, Biber's (2006) analysis of lexical bundles revealed that there were more of these bundles in classroom teaching than in other spoken registers. Biber (2006) associated the extremely high density of classroom teaching with the heavy reliance of this register on both written and spoken genres. Concerning discourse functions, classroom teaching included a higher proportions of stance bundles and discourse organizer bundles than textbooks. The findings revealed that lexical bundles functioned to organize discourse structures, signal transitions, transfer information, convey stance. However, the following sections will provide information about past related studies on lexical bundles with different research focus and different purposes as well as in different corpora.

## 2.3 Lexical Bundles and Written Academic Discourse

Research in the area of academic writing has been the interest of many scholars working in English for Specific purposes (ESP) and English for Academic Purposes (EAP). Studies of lexical bundles in a variety of written genres has received great attention from scholars for the last 20 years. The following discussion of related literature on the use of lexical bundles in academic written discourse is divided into three parts: 1) disciplinary study, 2) cross-disciplinary study, and 3) cross-linguistic study.

### 2.3.1 Lexical Bundles and Disciplinary Study

It seems that few studies have focused on the study of bundles within one single disciplinary area (Hyland, 2008b). The following studies basically focused on identifying lexical bundles in a certain discipline, looking for similarities and differences within the discipline. For example, Jalali, Rasekh and Rizi (2008) explored the frequency of occurrence, structural forms and discourse functions of lexical bundles. Their research focused on exploring possible variations within some written academic genres of one single discipline of applied linguistics. This study showed that there were relatively large differences between research articles and student genres in the types of bundles employed, and striking differences in terms of their frequency and function of 4-word lexical bundles. Farvardin, Afghari and Koosha (2012) looked at the way lexical bundles were used in research articles in physics in terms of the frequency of occurrence and distribution of different structural types. In their study, the research articles in the Physics Corpus consisted of 800 research articles which included 3 million words. Their results showed that most of the bundles used phrasal structures rather than clausal structures.

In the same way, Jalali (2013) explored possible similarities and/or differences between two students' genres in terms of the frequency, structures, and functions of lexical bundles. His data consisted of two corpora of master's theses and doctoral dissertations in one discipline of applied linguistics. The study revealed that there were striking differences in the frequency and function of these bundles as well as there was a large intradisciplinary difference between the two corpora in the range of bundles employed. Similarly, Jalali and Moini (2014) also examined the use of 4-word lexical bundles in only the introduction section of research articles in the field of medicine. The results of the study revealed that there were a total of 161 lexical bundle types that differed

in terms of structural forms and the most common structures found in the writers of medical research articles relied widely on noun phrases and phrasal bundles.

Qin (2014) investigated published research articles and research papers in the field of applied linguistics. He found that non-native graduate writers at the higher levels of study applied more structural forms in academic writing (e.g., noun phrases with post-modifier fragments) than those of lower levels. On the other hand, Grabowski (2015) focused on the exploration of lexical bundles in the pharmaceutical domain. The results showed that discourse functions of lexical bundles varied in different pharmaceutical texts. The association between the variability of lexical bundles pattern and their frequency could depend on genre or register. Studies by Marco (2000) and Mbodj-Diop (2016) also explored the frequency, functions and structures of lexical bundles in a corpus of 1 million of medical articles. Other studies (Weinstein, 2011; Shamsudin, Yusof & Raof, 2012; Huimin, 2014; AlHassan & Wood, 2015; Chen & Xiao, 2015; Güngör & Uysal, 2016; Li, 2016; Pičuljan, 2016; Tománková, 2016; Bychkovska & Lee, 2017; Kuo, 2017) examined lexical bundles in engineering fields, educational sciences, educational technology fields, applied linguistics and law. They found that there were differences with regard to some of the bundles analyzed in the academic texts written by native speakers and non-native speakers.

Few studies concentrated on pedagogical aspects of lexical bundles. For instance, Cortes (2006) explicitly taught lexical bundles to a class of history students who were taught intensive writing. She presented mini-lessons that were merged with the history class to cover the nature of lexical bundles, their functions, and activities targeting lexical bundles. The results of pre and post-analysis of students' writing did not show any significant improvements in terms of frequency of use of lexical bundles. However, she concluded that teaching lexical bundles explicitly could increase students' awareness of

these expressions and eventually lead to more appropriate academic writing. As can be seen, research has emphasized the extensive disciplinary differences between texts, but most of the existing research is found in the area of academic writing.

### 2.3.2 Lexical Bundles and Cross-disciplinary Studies

To explore disciplinary variations with regard to the use of lexical bundles in written texts, Hyland (2008a) analyzed forms, structures, functions of the frequently occurring 4-word lexical bundles in a corpus of 3.5 million words of published doctoral dissertations, Master's theses and research articles across four disciplines (biology, applied linguistics, electrical engineering and business studies). In this research, a lexical bundle had to occur with at least a frequency cut-off of 20 times in 1 million words and must appear in 10 % of texts. The frequencies and patterns of the bundles were compared across different disciplinary corpora. Through analyzing the forms, functions and structures of the lexical bundles, the results showed that academic written texts across disciplines displayed differences in bundle use. Over 50 per cent of the bundles in each discipline were unique and did not occur at all in any other disciplines. Only about one third of the items in each list were found in two other disciplines. The findings revealed that native academics used less formulaic expressions compared to postgraduate students that employed a higher number of these expressions. Hyland, also found variations across disciplines in the use of lexical bundles.

This study besides other similar studies indicate that the frequent and appropriate use of lexical bundles is a crucial part of fluent language production in academic environments. According to Hyland, (2008a, p.4), "helping to shape meanings in specific contexts and contributing to our sense of coherence in a text". Based on the functional framework established by Biber et al. (2004) and Biber and Barbieri (2007), Hyland redesigned a functional framework to fit into the analysis of bundles in its own context,

that is, for lexical bundles better suited for written research genres. The results showed that nearly half of the top 50 bundles in each list occurred in any other list, indicating that authors in different disciplines used different multi-word constructions to develop their arguments, persuade their readers, and establish their credibility.

Improving the understanding of multi-word sequences in the university context was always been a major concern among researchers and instructors to address this issue, Cortes (2004) examined the writing in published journal articles and the written production of native speakers' of English who were university students. Her corpus consisted of about 2 million words of two disciplines (biology and history). The results showed that expert writers used a higher number of lexical bundles than university students. Cortes's (2004) analysis was based on approximately one-million-word corpus for each field, and on approximately four hundred thousand words corpus for students' writing in each field. Among other analyses and comparisons in Cortes's (2004) work, the comparison between bundles in published disciplinary readings and students' writing revealed that most target bundles were never or rarely used, and that those that were found in students' writing did not align with similar functions of target bundles. Cortes (2004) suggested that exposing students to frequent lexical bundles was not sufficient to start using them in writing, and that more effort should be placed on getting students to notice such frequent lexical bundles in their respective fields. Similarly, Scott and Tribble (2006) examined the professional writings and the writing of MA students in British universities and concluded that MA students used more varied and sophisticated lexical bundles.

Further evidence that supports the fact the lexical bundles are building blocks of discourse is in a study by Strunkyt and Jurkunait (2008). They explored and compared the use of lexical bundles in the two major domains (natural sciences and humanities) in terms of their frequency of occurrence and their different functional and structural types

across disciplines. Each disciplinary division consisted of twenty research articles and they were extracted using WordSmith Tools computer software programme. The results showed that research articles in the field of humanities used a larger number of lexical bundles than those that were found in natural sciences. Also, Jalilifar, Ghoreishi & Roodband (2016) explored major disciplines as Arts and Humanities, Sciences, and Social Sciences. Likewise, several studies (Durrant & Mathews- Aydınlı, 2011; Maswana, Kanamaru & Tajino, 2013; Güngör, 2016; Güngör & Uysal, 2016) examined chunks in a variety of disciplines (e.g., anthropology, business, economics, mathematics, tourism, law, politics, and sociology). Byrd and Coxhead (2010) examined Academic Word List (AWL) and built a list of 21 of 4-word lexical bundles used in arts, commerce, law, and science through the analysis of a corpus of academic writing. Through this investigation, they were able to identify six key challenges in taking lexical bundles data into EAP classroom. Bal (2010) examined the four-word lexical bundles in published research articles from a total of 6 different academic disciplines which were written by Turkish researchers in international journals. Both functional and structural types were analyzed to discover any differences and/or similarities. He found out that over half of the lexical bundles in his study had not been identified before in the related research.

Another comprehensive study of lexical bundles in academic writing was conducted by Liu (2012). Using the academic writing sub-corpora of the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA), Liu identified 228 most frequently-used multi-word units, including lexical bundles, idioms and other expressions in academic writing in different disciplines. The results confirmed previous research findings in three respects. First, prepositional / noun phrases or fragments *(*e.g., *in terms of* and *a number of)* were the two major structural types of multi-word combinations in academic written texts, as was the case in Biber et al. (1999). Second, although phrasal verbs were rare in academic writing (Moon, 1998; Biber et al., 1999),

quite a few of the phrasal verbs occurred frequently in the texts under investigation, for example, *point out, bring about, carry out,* and *make up*. It also confirmed the findings in earlier research that lexical bundles were discipline-specific (Hyland, 2008a). Liu's study discovered some new findings that passive uses *(*e.g., *it has been suggested that)* exhibited a much lower frequency in academic writing, which was different from Biber et al.'s study (1999, p. 1019-1021). Some multi-word constructions that were reported as being frequently used in previous research (Biber et al., 1999; Carter & McCarthy, 2006; Simpson-Vlach & Ellis, 2010) were observed with a lower frequency than those found by other researchers.

Durrant (2013) studied students' essay writing in the corpus of British Academic Written English (BAWE). He examined at a corpus of a total of 1,558 texts in 24 different disciplines. He used Hyland (2008)'s framework and made modifications to it. Four major disciplines were recognized, namely, life sciences, humanities and social sciences, commerce, and science and technology. Durrant (2013) explored the disciplinary variations in the use of 4-word lexical bundles in university students' writing. This study was different than previous studies in that the disciplinary groupings were not assumed at the beginning of the research, but rather at the start of the analysis of the corpus. The variation and degrees of differences and similarities between individual writers were demonstrated in the form of a visual map. In their research, they employed a qualitative analysis of lexical bundles which distinguished between soft and hard disciplines in order to characterize the discourse functions.

Durrant (2017) also highlighted the disciplinary variations in the writing of university students. He identified some differences between disciplinary groupings (science and technology, humanities and social sciences). The results showed that research-oriented bundles were the focus of the disciplines of science and technology; however, bundles in

humanities and social sciences focused on abstract concepts. Procedure bundles were more frequent in the fields of science and technology. Durrant discovered that humanities and social sciences consisted a larger number of stance bundles; however, the text-oriented bundles were more frequent in science and technology.

Likewise, Kwary, Ratri and Artha (2017) examined the frequency of lexical bundles, their functional categorizations and their structural types in a corpus of 2,937,431 words. The corpus consisted of a total of 400 research articles that were distributed in four academic disciplines: life sciences, social sciences, health sciences and physical sciences. The results revealed that physical sciences used the larger number of lexical bundles. Shared bundles were found between academic disciplines in physical sciences and health sciences, whereas, a few were found between the social sciences and health sciences. On the other hand, Esfandiari and Moein (2016) explored the frequent four-word lexical bundles and their functions in a corpus of 4,652,444 research articles in two disciplines food science and technology. They used Hyland's (2008b) functional taxonomy to analyse the functions of the lexical bundles. Their results of frequency and range showed that text-oriented lexical bundles were the highest in the whole corpus.

Johnston (2017) compiled four corpora of writing in literature and applied linguistics, representing professional and learner writing in each field. She identified four-word lexical bundles and looked at frequency, structure, and functions of these bundles. The results revealed that expert writers in applied linguistics and literature used bundles with different frequency, displayed different choices of lexical items to fill structural bundles, and used functional bundles differently. Learners in each field displayed differences in their use of bundles as compared to the professionals' use. Learners in applied linguistics used more types and tokens of bundles overall, while learners in literature used fewer.

Both groups of learners relied more on repetitive use of certain bundles than did the professionals.

### 2.3.3 Lexical Bundles and Cross-linguistic Studies

The discussion moves to lexical bundles in the context of cross-linguistic studies with a brief review on native and non-native language acquisition, for example in languages like Spanish (e.g., Butler, 1998; Tracy-Ventura, Cortes & Biber, 2007; Cortes, 2008), Korean (Kim, 2009) and Chinese (e.g., Wei, 2007; Xu, 2007; Ma, 2009).

A study by Ädel and Erman (2012) compared the use of 4-word lexical bundles in the writing of English native students with the written production of advanced learners who were native speakers of Swedish. For their corpora, Ädel and Erman (2012) used a 247-thousand-word corpus and an 863- thousand-word corpus for English native speakers and non-native speakers. Results showed that native students used a wider number of lexical bundles types than that of non-native learners. Non-native speakers, in Ädel and Erman (2012)'s study, used not only less lexical bundles (115 bundles compared to 185 bundles), but also they used them with less variation. Moreover, non-native speakers displayed signs of register difficulties shown by the lexical items used within lexical bundles (Ädel & Erman, 2012).

In another study, Cortes (2008) compared published articles of history in English and in Spanish. Research articles of history articles were written in American English and Argentinean Spanish. In both corpora, lexical bundles were identified and analyzed in terms of both functions and structures. The results by Cortes showed that there was a difference in the use of lexical bundles as well as a certain level of agreement in the bundles used in each language. Kim (2009) explored other languages other than English like Korean. The large corpus consisted of conversation and academic written texts. The

results of her study revealed that Korean language depends on lexical bundles as they are considered as important expressions that function as discourse organizers. On the other hand, Allen (2009) investigated the use of 4-word lexical bundles in a number of science research articles written in English by first-year undergraduate science majors at the University of Tokyo. Allen identified a total of 144 lexical bundles and analyzed the functions of the frequently occurring lexical bundles. He found that the research-oriented bundles were the most frequent. Further analysis of the results revealed that certain lexical bundles are suitable for designing learning activities, using concordances.

Some studies that focused on lexical bundles showed that the knowledge of these expressions led to a high proficiency in the language more than the knowledge of individual words. For instance, Vidakovic and Barker (2010) found that lower proficiency leaners depended more on single words rather than lexical bundles. The study by Chen and Baker (2010) compared lexical bundles in two corpora of published academic writing of native English students with the research articles by EFL Chinese students. Their results showed that non-native learners' writing showed a smaller number of lexical bundles. Meanwhile, both native students and L2 students underused certain high-frequency lexical bundles identified in published articles, such as *in the context of*, while some expressions such as *all over the world* were overused in the L2 student corpus. Pang (2010) examined lexical bundles in order to show that these expressions have an important role building coherence in academic writing. They made an attempt to explore the different strategies that L2 learners follow when using these expressions. Also, Rafiee et al., (2011) explored the different structural types in different corpora of English newspapers which was edited by native speakers and non-native speakers.

Similarly, Karabacak and Qin (2013) examined the use of lexical bundles in term papers written by Turkish, American and Chinese first and second university students.

Their comparison showed that American students used a larger number of these expressions compared to those used in Turkish and Chinese students. On the other hand, Salazar (2010) analyzed the use of lexical bundles in varieties of English. They extracted lexical bundles in two corpora of medical research articles taken from Philippine English language journal and another from the British medical journal. Her quantitative analysis showed a lower amount of these expressions in the corpus of Philippine English compared to British, while her qualitative findings uncovered certain functional and structural differences between the expressions used in two corpora. Salazar (2013) derived lexical bundles to compare their use in two corpora. One sub-corpus consisted of a number of published articles in biology written by native and non-native English speakers, while the other one contained research articles from biochemistry journals also written by native and non-native English speakers.

In another study, Beng and Keong (2014) analyzed the use of 4-word lexical bundles in reading passages in the Malaysian University English Test (MUET). The objective of their study was three-fold: (1) to investigate the most frequent lexical bundles in the corpus, (2) to classify the structural types of lexical bundles, and (3) to compare the grammatical forms in (Arts and Science) texts. Beng and Keong identified 730 lexical bundles including two-, three-, and four-word lexical bundles. In addition, results revealed that, to convey the intended meaning, authors seek the help of certain patterns of language mainly based on the context of the reading passages. In the same way, Pérez-Llantada (2014) explored lexical bundles in L2 English published writing and L1 English published writing in research articles in twelve different disciplines. Findings showed that the L2 English writers used formulaic language not fully native-like.

Concerning proficiency levels, Staples, Egbert, Biber, and McClair's (2013) used the (TOFEL iBT) writing section to study the use of lexical bundles. The results generally

revealed that the students with lower proficiency levels made use of a larger amount of lexical bundles than those students that had a higher proficiency level. However, there were not so many differences regarding the discourse functions across different proficiency levels. Chen and Baker (2016) also derived lexical bundles from a corpora of L2 written essays. Their results revealed that lower level of proficiency students made use of more structural types that represented verb lexical bundles/clausal bundles, whereas noun phrase and prepositional phrase lexical bundles were extensively found in the writing of advance. Pan, Reppen, and Biber (2016) compared two corpora of research articles from Telecommunications journals written by native-speakers of English with non-native speakers of English. They found that the both groups of writers employed different discoursal functions and different structural types of lexical bundles.

Several other researchers have identified lexical bundles to create lists of expressions that could have more pedagogical implications (Simpson-Vlach & Ellis, 2010; Wood & Appel, 2014). For example, Wood and Appel explored the most frequent four-word lexical bundles in first year engineering and business university textbooks to create a list of these expressions that could be used pedagogically.

Findings of all these studies found that lexical bundles differed from one discipline to another and emphasized the need to further explore these lexical bundles in various academic disciplines.

## 2.4 Lexical Bundles and Spoken Academic Discourse

Limited number of spoken corpora is available for researchers due to the fact that the transcription of a spoken corpus into written format is a very demanding and time-consuming process (Biber et al., 2004; Nesi & Basturkmen, 2006; Biber & Barbieri, 2007;

Neely & Cortes, 2009; Csomay, 2013; Hernández, 2013; Kashiha & Chan, 2013, 2014a, 2014b; Kashiha, 2015).

The importance of lexical bundles in creating fluency in language production has been addressed by many researchers but not many studies have concentrated on the distribution of these expressions in disciplinary lectures. Nesi and Basturkmen (2006) did not explicitly investigate the use of lexical bundles with regard to disciplinary variation. The study aimed to contribute to the description of disciplinary lectures by comparing frequency distribution of 4-word lexical bundles to see the way in which disciplinary lectures used the combination. Nesi and Basturkmen (2006) examined a total number of 160 monologic academic lectures from two online corpora: BASE and MICASE (Michigan Corpus of Academic Spoken English). The functions of lexical bundles in the lectures were examined in order to investigate their signaling role. Results showed that there was a great dependency found between the frequent use of these bundles and the organization of academic lectures. Therefore, the findings indicated that there was a singling role of these expressions in discourse. They concluded that it is necessary to raise students' awareness of the use of lexical bundles.

Simpson (2004) explored the MICASE which included 1.7 million words. He investigated the monological and interactive discourse between professors and students in academic settings (native and non-native speakers). By studying a list of 224 expressions that were structurally coherent and idiomatically complete, Simpson reported that interactive speech had more than twice as many bundles as were found in monologues. Some bundles *(*e.g., *and so on)* were more frequently used by professors, while students used some expressions like *something like that.* Likewise, Neely and Cortes (2009) used MICASE online corpus to examine the spoken language of lecturers and students to identify the functions of five lexical bundles *(a little bit about, I want you*

*to, a little bit of, I would like you* and *if you look at)* which were generally used to introduce a new topic and to organize discourse in spoken academic discourse. They suggested that lexical bundles should be taught to learners in a rather natural way. The findings of their study were consequently used to develop and design an EAP listening curriculum.

Two more studies on academic spoken discourse by Herbel-Eisenmann and Wagner (2010) and Herbel-Eisenmann, Wagner and Cortes (2010), closely observed the use of lexical bundles in two corpora: one consisting of 148 transcripts in secondary mathematics classrooms, the other from conversations. The conversation corpus was taken from the Longman Grammar of Spoken and Written English (LGSWE) corpus. Researchers found that stance bundles were found to occur more frequently in secondary mathematics classrooms. They served to express personal values, attitudes and feelings. Mathematics teachers used stance bundles in the opening or closing of classroom teaching sessions. They highlighted that stance bundles helped students and teachers to engage in conversations in classrooms and to express authority.

On the other hand, Csomay (2013) studied the distributional patterns of the functions of lexical bundle reported in previous studies (Biber et al., 2004) and explored how they were related to discourse structure. A total number of eighty four lexical bundles and their functions (Biber et al., 2004) were examined in nearly 1,176 units of discourse extracted from the introduction part of 196 university lectures taken from two different corpora: MICASE corpus and TOEFL 2000 Spoken and Written Academic Language Corpus (Simpson & Swales, 2001). The findings revealed that there was an obvious link between the functions of the lexical bundle and the communicative purposes in a discourse structure. However, discourse organizing bundles were least prominent. Findings indicated that functions of lexical bundle linked well to the functions in discourse

structure found through linguistic variation (Csomay, 2005) and suggested that there was a great connection between lexis and grammar on the discourse level. Another research was conducted by Hernández (2013) in which he examined the frequency forms, discourse functions and structural types of four-word bundles in three corpora of spoken English, two of them were non-native speakers of English and the other one was a sub-corpus of native speakers of English, corresponding to university students in their first year of an English Studies degree and to the same students after two years of university instruction. The study focused on three major characteristics: the overall distribution of bundles, their typical structures and their functions. The findings showed significant variations in the types of lexical bundles used by students who were native speakers and non-native speakers, as well as in their structure and function. The results supported the idea that lexical bundles are important components in oral discourse.

Also, another important area of investigation is the exploration of lexical bundles with regard to genre and disciplinary variation in two corpora of spoken academic ELF, namely lectures and seminars. Wang (2017) explored the influence of disciplinary variation and genre on the use of 4-word lexical bundles in a corpus of transcribed spoken academic lingua franca English (ELFA). The results demonstrated that these expressions provided valuable indications on disciplinary variations and genre. Wang investigated the use of 4-word lexical bundles in terms of their frequency of occurrence, functions, and structural types in different genres: (seminars and academic lectures) and across 3 disciplines (Social Sciences, Natural Sciences, and Medicine). The results revealed that discipline and genre were two essential elements that were useful indicators in understanding the communication in academic ELF. Similarly, Wang (2018) used a part of lecture subset used in Wang (2017). Their results provided further evidence for disciplinary differences and variability in the use of formulaic language to signal discourse organization by lecturers in academic ELF settings.

However, there is increasing evidence that subject discipline plays an important role in academic speech (e.g., Kashiha & Chan, 2013; 2014a, 2014b; Kashiha, 2015; Wang, 2017). Both Kashiha and Chan (2014) and Wang (2017), for instance, identified important differences in the frequency forms, structural types and discourse functions of lexical bundles in different academic disciplines. These studies investigated the use of lexical bundles in different academic disciplines. Thus, most previous studies pointed to the considerable variation of bundle use in relation to other linguistic aspects in university lectures in different academic disciplines, whereas, the discipline-specific use of those bundles had not been considered. However, the analysis of lexical bundles in specific disciplines is an important theme or aspect that should be given greater attention in lexical bundles research. For example, one of these studies is the research done by Kashiha and Chan (2013) who examined how 4-word lexical bundles were used in two corpora of twenty four university lectures on soft and hard sciences that were extracted from the British Academic Spoken English (BASE) corpus. This was done in order to identify the most frequently occurring lexical bundles to highlight the differences and similarities in terms of their frequency forms, communicative functions and structural types. They also explored their distribution of the extracted bundles in the introduction, body and closing sections of university lectures. The findings showed some differences between the two disciplinary divisions. Lecturers in each sub-corpus seemed to use lexical bundles with various discourse functions and structural types in order to convey their messages.

In their subsequent studies, Kashiha and Chan (2014a) carried out a contrastive corpus-based study to examine the structural types of the most frequent 4-word lexical bundles in two corpora of a total number of eight academic lectures in two different disciplines: Politics and Chemistry that were taken from BASE corpus. They highlighted the differences and similarities of disciplinary variations in the frequency and forms of the structures of the lexical bundles. They adopted Biber et al.'s (2004) structural taxonomies

to classify lexical bundles. The results showed that the most common lexical bundle types in academic lectures of the two disciplines were noun phrases and prepositional phrase fragments. The findings revealed that academic lecturers appeared to use a number of different structural types of lexical bundles to deliver their disciplinary messages.

Similarly, Kashiha and Chan (2014b) as in their previous study, used the same lectures (8 lectures from Politics and Chemistry) taken from BASE corpus to investigate the discoursal functions of lexical bundles that were found in university lecturers. Their study aimed to compare the most common 4-word lexical bundles of university lectures in the two different disciplines in order to categorize these expressions functionally. Kashiha and Chan adopted the framework proposed by Biber et al. (2004). They classified the identified lexical bundles according to their functional categorization in order to explore their communicative purposes. Also they examined expressions conveyed in the two groups of lectures to look at differences and similarities within the two disciplines. Their findings highlighted some striking variations of these expressions which were found between the two disciplines in terms of communicative functions of the lexical bundles. Results revealed that academic lectures depended extensively on the use of certain expressions to fulfill communicative purposes.

Similarly, Kashiha (2015) conducted a comparative study which was run on a 1 million word corpus of 120 academic lectures taken from the BASE corpus. Kashiha investigated and compared the frequency, structure, and discourse function of the frequently occurring 4-word lexical bundles in academic lectures across three broad disciplinary groups, namely Social Sciences, Physical Sciences and Life Sciences. The functional and structural taxonomies proposed by Biber et al (2004) were used as analytical framework to group lexical bundles in terms of their grammatical types and the discourse functions they serve. Findings revealed some variations between the three corpora in relation to the

distributional patterns of the target bundle. The three groups of lecturers also showed different tendencies towards the selection of grammatical types to form these expressions and the communicative functions that the these bundles carried out in university lectures.

As seen in the studies above, there are some differences and similarities in the use of lexical bundles between written discourse and spoken discourse. The use of lexical bundles in many studies differ between spoken and written discourse. As shown in Biber et al. (2004), lexical bundles as building blocks of spoken and written discourse, occur more frequently in spoken discourse (classroom teaching and conversations) than in written discourse (academic textbooks and academic prose). This may indicate that the speaker need to make more effort to inform the audience of the direction of the discourse. As an example of this, the majority of bundles in academic writing tend to be based on noun phrases (*the results of the, the fact that the,* and *a large number of*) and prepositional phrases (*in the case of, at the end of,* and *on the basis of*), while the majority of lexical bundles in spoken discourse are based on verb phrases (Biber et al., 2004).

The findings of Kashihas' studies revealed that lexical bundles were frequently used in academic lectures and accounted for a large proportion of lecturers' speech. In addition, various disciplines differed in how they used lexical bundles to engage listeners, expressed their stances and provided referential information for precise presentation of disciplinary content, and organized the flow of their speech. Lecturers in each field of study relied on a variety of forms and functions to convince the listeners, get their disciplinary messages across and communicate their arguments. This indicates that disciplinary variation could impinge on the choices of the lexical bundles.

On the whole, findings of previous studies revealed that hard disciplines (e.g., science and technology) offered a larger number of lexical bundle use compared to that of soft disciplines (e.g., humanities and social sciences). There were also some notable

similarities and differences in the structural types of the bundles used. Noun and prepositional phrase fragments were the most commonly used structures in soft science corpus, while these structures along with dependent clause fragments were the most prevalent in hard sciences. With regards to the functional perspectives of the lexical bundles used in the lectures, lecturers' in hard sciences mainly use stance bundles to directly guide listeners through the content of the lecture while lecturers in soft sciences make use of stance bundles to cautiously construct their arguments. Findings reported the higher concentration of referential expressions in the soft science lectures, whereas lecturers in hard sciences showed a greater tendency towards the use of discourse organizers.

## 2.5 Discussion on Reviewed Studies

The majority of research studies have concentrated on the investigation of lexical bundles in a variety of genres, namely academic prose and conversations (Biber et al., 1999), classroom teaching and textbooks (Biber et al., 2004); research articles, doctoral dissertations and Master's theses (Hyland, 2008); writing-intensive history (Cortes, 2006); and academic lectures (Nesi & Basturkmen, 2006). These studies on lexical bundles have targeted academic written register rather than spoken. For example, Hyland (2008) examined the use of lexical bundles in the Master's theses and doctoral dissertations. Also Adel and Erman (2012) studied lexical bundles in academic writing of native speakers and non-native speakers, and Cortes (2006) explicitly taught lexical bundles to students of university in a writing intensive history class and investigated its effects.

Concerning spoken registers, little attention has been given to studies that have specifically compared the use of lexical bundles in university lectures of different disciplinary divisions. Nesi and Basturkmen (2006) investigated the lexical bundles with

regard to the cohesive roles they play in a total number of 160 university lectures. In some studies, lexical bundles are analyzed quantitatively, with the results reporting only the frequency of the identified bundles without giving a comprehensive explanation on the reasons behind the bundle use (Rafiee et al., 2011).

Therefore, with the above concerns in mind, the present research is motivated by a lack of empirical research to analyze and compare the use of four-word lexical bundles in university lectures of different disciplinary divisions. In other words, a need is felt to explore the variations across disciplinary divisions in the use of lexical bundles in lectures to find out how different lecturers organize their discourse to get their message across.

## 2.6 Summary of the Chapter

From the related literature, it seems that research studies on the language use of lexical bundles in academic lectures are few with regard to the context of their use in broad disciplinary divisions. Therefore, it is noteworthy to explore the lexical bundles used in disciplinary domains that include a greater number of disciplines in larger corpora to reveal types, functions and structures.

This study extends the research on lexical bundles by examining 4-word lexical bundles in two corpora of academic lectures delivered by native speakers with regards to disciplinary variations. The analysis of lexical bundles in both corpora is compared by exploring the frequencies of these expressions in both groups of disciplines. Also, the functional and structural distributions of four-word lexical bundles will be analyzed in each discipline.

# CHAPTER 3: METHODOLOGY

## 3.1 Introduction

This chapter describes the steps taken to conduct this study. In the first section, the collection of the corpus created for the purpose of this study (a corpus of academic lectures) will be introduced. In the second section, the concordancing program used to facilitate the search for lexical bundles in the corpus will be described, and in the last section the taxonomies used for structural and functional analysis of the identified lexical bundles will be discussed.

## 3.2 Research Design

The present study is a corpus-based study which includes both quantitative and qualitative approaches as shown in Figure 3.1. In this study, the quantitative analyses concerns the frequency of lexical bundles and their patterns of use in academic lectures of Arts and Humanities, and Life and Medical Sciences disciplines. This study also discusses the structures and discourse functions of the identified lexical bundles in each disciplinary division as well as explanations of their language use. These are presented through the qualitative analysis. It is believed that using a combination of both research designs provides a better picture of dealing with research problems rather than that of individual quantitative and qualitative analyses.

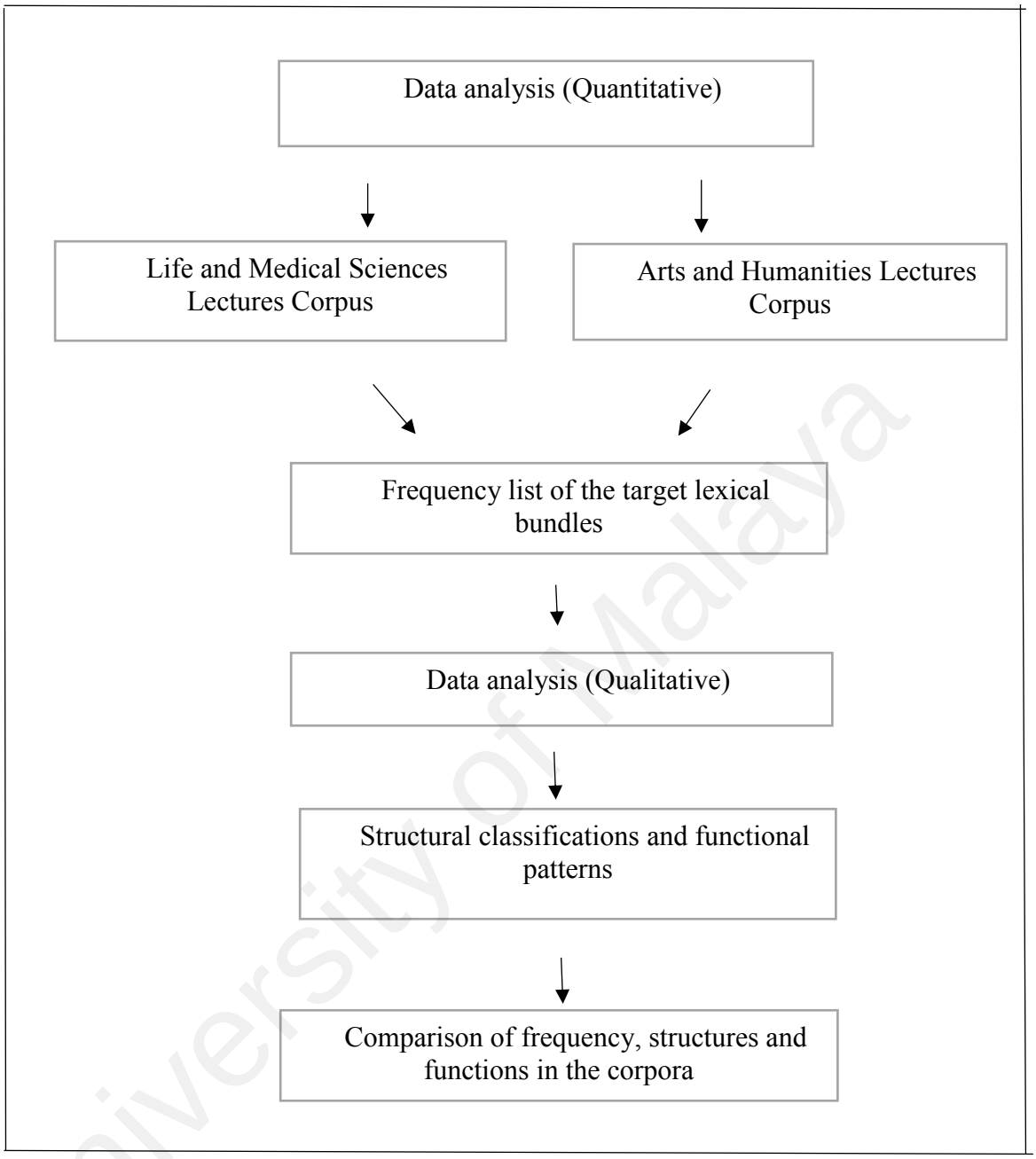


**Figure 3.1: Conceptual Framework of the Design**

## 3.3 The corpus

According to Paltridge (2006), a corpus is a collection of written and spoken texts representing particular areas of language use. Paltridge (2006) also suggests that there are two types of corpora: a general corpus and a specialized corpus. In a general corpus, the

language use is represented by and large leading to more generalizable results. On the other hand, a specialized corpus represents a particular aspect of language use. This type of corpus refers to a set of texts representing a specific genre such as academic lectures or research articles, collected to investigate the use of particular linguistic features. In relation to the size of the corpus, Biber (2006) argues that it depends on the aim of the study; however a corpus must be large enough to sufficiently symbolize the occurrence of the linguistic feature being analyzed.

This study was carried out on a corpus of 80 academic lectures recordings (40 from Arts and Humanities and 40 from Life and Medical Sciences). The recordings transcriptions used in this study come from the British Academic Spoken English (BASE) corpus. The corpus development was assisted by funding from universities of Warwick and Reading, United Kingdom. The rationale behind using BASE for this present study is because it is the largest online corpus available on academic lectures and it has been designed to represent as well as make available a variety of linguistic characteristics of spoken academic discourse (Kashiha, 2015). The data selected for this study are of monolingual and transactional academic university lectures across two broad disciplinary groupings, Arts and Humanities (AH) and Life and Medical Sciences (LS). Each academic division comprises 40 lectures across a number of sub-disciplinary fields of research. Table 3.1 gives the details about the corpus used in this study.

**Table 3.1: information about the academic lectures used in the study**

| Corpus | No. of lectures | Word count |
|---|---|---|
| **Arts and Humanities** | 40 | 306,652 |
| **Life and Medical Sciences** | 40 | 288,226 |
| **Total** | 80 | 594,878 |

## 3.4 Data Collection Procedure

The first step is to download the 80 lectures from BASE corpus across the two broad disciplinary groupings of Arts and Humanities, and Life and Medical Sciences. BASE is an online academic corpus consisting of 160 lectures and 40 seminars recorded across four broad fields of studies (Arts and Humanities, Life and Medical Sciences, Social Sciences and Physical Sciences). Social sciences and Physical sciences are not included in this study. The rationale behind choosing the two divisions (Arts and Humanities, and Life and Medical Sciences) is because Arts and Humanities is a broad disciplinary group that includes a number of non-scientific academic disciplines and a great number of these disciplines fall under soft sciences (Becher, 1987, 1994). On the other hand, Life and Medical Sciences include scientific disciplines and most of these fields of studies fall under applied and pure sciences (Becher, 1987, 1994). The non-scientific academic division of Arts and Humanities deal with human aspects and it includes academic disciplines such as Caribbean studies, Classics, Comparative American studies, English, Film and Television Studies, French, History, History of Arts, Philosophy and Typography. However, the scientific academic division of Life and Medical sciences deal with the study of human life and organisms and it includes the following academic disciplines: Agricultural Botany, Zoology, Animal and Microbial sciences, Biological sciences, Mathematics, Medicine, Statistics and Plant Sciences.

The main focus of this present study is to explore and compare the lexical bundles used in two different broad disciplinary groupings, namely, Arts and Humanities, and Life and Medical Sciences. These two academic divisions were selected because they were less explored in the literature compared to the other two academic divisions such as social sciences and physical sciences (Kashiha, 2015). Both the Arts and Humanities, and Life and Medical Sciences lectures (compared to the size of the other two academic divisions)

are nearly similar in word counts which makes both corpora comparable in size. The rationale for choosing BASE is because BASE compared to other available corpora like MICASE (Michigan Corpus of Academic Spoken English), is the only corpus which holds this amount of transactional academic lectures across the two academic divisions.

After all the transcripts had been collected, they were then saved into Microsoft word files in order to be labelled individually according to each division. Finally, the Microsoft word files were converted into text files (stored in the form of txt. Files) so as to be readable by the computer software program AntConc 3.4,0 (Anthony, 2014) for data analysis.

## 3.5 Analytical Framework

The following section describes the frequency as well as structure and function analysis of lexical bundles across the two broad disciplinary groupings.

### 3.5.1 Lexical Bundle Identification

Frequency cut-off was normalized at 20 times per 1 million words for the programme. A lexical bundle has to occur in at least 4 lectures for it to be considered in the study. These methods of normalization have been broadly used in different corpora size to compare lexical bundles (Biber et al., 2004).

### 3.5.2 Data Analysis Procedure

The retrieved bundles were checked manually to exclude those bundles that were too speech-dependent and discipline-dependent such as proper nouns. The rationale for excluding these bundles results from the attempt to guard against the subjective influence which may be introduced by the topic of the lectures. Finally, only those 4-word lexical bundles which met all the above mentioned criteria and norms were selected and listed

for analysis and these were used as a basis for comparing the lexical bundles used in the lectures of the two broad disciplines. Once the lists of identified 4-word lexical bundles in both sub-corpora were prepared, they were checked manually to find those bundles which were shared among both academic divisions. Next the distribution of the target bundles in each division were calculated individually to come up with the number of different bundle (types) as well as individual cases (tokens). Then their frequency and patterns of use in both groups were compared to see which group of lectures relied more on the use of lexical bundles in presenting disciplinary materials. After comparing their frequency of use, the bundles were categorized structurally according to their grammatical focus using Biber et al. (2004)'s framework.

Finally, the identified bundles were also categorized according to their functions or contextual meanings. This study adopts Biber et al. (2004)'s framework as it is found to be suitable for research on spoken registers as attested by other researchers as well (Biber & Barbieri, 2007; Kashiha & Chan, 2013, 2014; Kashiha, 2015; Nesi & Basturkmen, 2006). To this aim, all the retrieved bundles were analyzed manually in the context of both academic divisions to decide on the type of communicative functions they were conveying or trying to convey.

Lexical bundles as target linguistic functions in the study were identified first and then analyzed in their context in each sub-corpus to discover the discourse communicative functions that they conveyed. It is necessary to explain that all the examples (tokens) of each lexical bundles type were analyzed individually in their context in order to establish a better understanding of the communicative functions of each expressions. If a lexical bundle had more than one function, the one which had been used the most was considered as the main function and the others were discussed as its secondary function and might vary across the sub-corpora.

The quantitative analysis of the present study included the comparison of both discipline division in terms of the distribution of the main functions of the lexical bundles. For example, what similarities and differences were found in the lexical bundles used in lectures in the Arts and Humanities discipline, compared to the lectures in the Life and Medical Sciences?

### 3.5.3 Analytical Framework of Lexical Bundles

To analyze the structure or grammatical type of lexical bundles, Biber et al.,'s (2004) classification is used. In order to analyze the function or meaning of lexical bundles, two frameworks were available, one by Hyland (2008) and the other, Biber et al., (2004). The rationale behind using Biber et al.,'s framework is because their framework is designed for research in both spoken and written registers, while Hyland's taxonomy is only designed to specifically reflect the concerns of research writing. It classifies lexical bundles used in written discourse such as research papers, doctoral dissertations and Master's theses. Hyland (2008, p.13) commented that Biber et al.'s classification is based on a much broader corpus of both spoken and written registers including class conversation, textbooks, course books, service counters and so on, and this seemed to have yielded for more personal, referential and directive bundles than many more research-focused genres.

Hence, the rationale for the selection of Biber et al. (2004)'s taxonomy is that they have made a distinct classification for the analysis of structural and grammatical features of lexical bundles found in spoken discourse. Therefore, this study adopts Biber et al. (2004)'s framework as it is found to be suitable for the research on spoken registers. Table 3.2 shows the structural classification of lexical bundles following Biber et al., (2004).

**Table 3.2: Structural classifications of lexical bundles (Biber, *et al*., 2004, p.381)**

| Structural categories | Sub-categories | Example bundles |
|---|---|---|
| **1. Lexical bundles that incorporate verb phrase fragments** | 1a. (connector+) 1st/2nd person pronoun + VP fragment | *I'm not going to* |
| | 1b. (connector+) 3rd person pronoun + VP fragment | *and this is a* |
| | 1c. discourse marker + VP fragment | *I mean I don't* |
| | 1d. Verb phrase (with non-passive verb) | *have a lot of* |
| | 1e. Verb phrase (with passive verb) | *is based on the* |
| | 1f. yes-no question fragments | *are you going to* |
| | 1g. WH- question fragments | *what do you think* |
| **2. Lexical bundles that incorporate dependent clause fragments** | 2a. 1st/2nd person pronoun + dependent clause fragment | *I want you to* |
| | 2b. WH-clause fragments | *when we get to* |
| | 2c. If-clause fragments | *if we look at* |
| | 2d. (verb+adjective) to-clause fragment | *to be able to* |
| | 2e. That-clause fragment | *that this is a* |
| **3. Lexical bundles that incorporate noun phrase and prepositional phrase fragments** | 3a. (connectors+) Noun phrase with of-phrase fragment | *one of the things* |
| | 3b. Noun phrase with other post-modifier fragment | *the way in which* |
| | 3c. Other noun phrase expressions | *a little bit more* |
| | 3d. Prepositional phrase expressions | *at the end of* |
| | 3e. Comparative expressions | *as well as the* |

Biber et al.,'s (2004) structural taxonomy includes three main grammatical types for lexical bundles: 1) lexical bundles that incorporate verb phrase fragments such as *is based on the*; 2) lexical bundles that incorporate dependent clause fragments like *I want you to*; and 3) lexical bundles that incorporate noun phrase and prepositional phrase fragments such as *at the end of* (Biber et al., 2004, p.381). Each main structural type has several sub-structures which are listed in Table 3.2. The first main structural category "lexical bundles that incorporate verb phrase fragments" is divided into seven types. For example, types 1a and 1b begin with a subject pronoun followed by a verb phrase (e.g., *I'm going to give* and *it's going to be*). Type 1c bundles begin with the discourse markers followed by a

verb phrase (e.g., *you know this is* and *I think you know*). Type 1d and 1e begin with a verb phrase (e.g., *are going to be*), while types 1f and 1g are question fragments (e.g., *what's going to happen* and *what do you think*).

The second main structural category "lexical bundles that incorporate dependent clause fragments" is divided into five sub-categories. For example, Type 2a bundles begin with a subjective pronoun followed by verb phrase (e.g., *I want you to*). Type 2b begins with a WH –word introducing a dependent clause. Other types such as Type 2c begins with a main clause followed by a complementizer (e.g., *if we look at*) or as Type 2d lexical bundles are dependent clause fragments that begin with a complementizer or subordinator (e.g., *to be able to*). The lexical bundles that belong to structural categories (1) and (2) are clausal while the structural category "noun phrase+ prepositional phrase fragments" are phrasal. Type 3a begins with a noun phrase and followed by of-phrase fragments (e.g., *a little bit of* and *at the start of*). Type 3b noun phrase followed with other post-modifier fragments (e.g., *those of you who* and *the way in which*). Type 3c simply begins with a noun phrase. Type 3d consists of prepositional phrase components with embedded modifier (that) as in (e.g., *of the things that*), while Type 3e incorporate comparative expressions.

**Table 3.3: Discourse functions of lexical bundles (Biber, et al., 2004, p.384-388)**

| Functional categories | Sub-categories | Sample bundles |
|---|---|---|
| **1. Stance expressions** | A. Epistemic stance<br>Personal<br>Impersonal | <br>*I don't know if,*<br>*the fact that the* |
| | B. Attitudinal/modality stance | |
| | B1)Desire | *if you want to,* |
| | B2)Obligation/ directive<br>Personal<br>Impersonal | <br>*I want you to, it is*<br>*important to,* |
| | B3) Intention/ Prediction<br>Personal<br>Impersonal | <br>*I'm not going to,*<br>*it's going to be,* |
| | B4) Ability<br>Personal<br>Impersonal | <br>*to be able to, to*<br>*can be used to,* |
| **2. Discourse organizers** | A. Topic introduction/focus | *what do you think* |
| | B. Topic elaboration/ clarification | *on the other hand,* |
| **3. Referential expressions** | A. Identification/ focus | *that's one of the,* |
| | B. Imprecision | *or something like that* |
| | C. Specification of attributes | |
| | C1) Quantity specification | *there's a lot of* |
| | C2) Tangible framing | *the size of the* |
| | C3) Intangible framing | *the nature of the* |
| | D. Time/ Place/ Text reference | |
| | D1) Place reference | *the united states and* |
| | D2) Time reference | *at the same time* |
| | D3) Text-deixis | *as shown in figure* |
| | D4) Multi-functional reference | *at the end of* |
| **4. Special Conversational Functions** | A. Politeness | *Thank you very much* |
| | B. Simple inquiry | *What are you doing* |
| | C. Reporting | *I said to him/her* |

The identified bundles were also categorized according to their function or contextual meanings. This study adopts Biber et al. (2004)'s framework and the retrieved bundles were analyzed manually in the context of both academic divisions to decide on the type of their communicative functions. According to Biber et al., (2004),'s taxonomy, lexical bundles are distributed across the three main functional categories. Each of these

functional category involves several sub-categories linked to more specific discourse functions and meanings:

(1) Stance bundles are defined as the "overt expression of an author's or speaker's attitudes, feelings, judgments, or commitment concerning the message" (e.g., *I don't know if* and *the fact that the*) (p. 386). Stance bundles convey two major functions: epistemic and attitude/modality. Epistemic stance bundles comment on the knowledge status of the information: certain, uncertain, or probable/possible (e.g. *I don't know what*, *I don't think so*, and *the fact that the*). There are two sub-categories of epistemic stance bundles: Personal and impersonal. Personal epistemic bundles can express certainty and uncertainty but most of them express only uncertainty (e.g., *I don't know what, I don't think so,* and *I think it was*) are examples of personal epistemic bundles. Impersonal epistemic bundles express certainty (e.g., *are more likely to* and *the fact that the*).

Attitudinal/ modality stance bundles express speakers' attitudes towards the actions/events described in the following proposition (e.g. *I don't want to* and *i'm not going to*). The attitudinal/ modality bundles are divided into four functions: desire that expresses only personal attitude (e.g. *I don't want to*). Personal stance bundles are only included in the desire bundles. These bundles frame self-motivated desires and wishes or ask about another participant's desires by using the personal pronouns (*I*), for example, (*I would like to*, and *what I want to do*). Impersonal stance bundles express similar meanings without being attributed directly to an individual (e.g. *it is possible to*).

Obligation/directive are also personal stance bundles (e.g., *I want you to, you have to do,* and *you need to know*). These bundles have a second person pronoun (*you*) instead of (*I*). A few of these bundles are impersonal with no pronoun such as *it is important to*. Intention/ prediction are mostly personal (e.g., *I'm not going to* and *it's going to be*).

They express the speaker's/writer's future intentions. Some bundles express ability (e.g., *to be able to* and *to come up with*).

(2) Discourse organizers indicate the relationships between given and coming discourse. They have two major functions: topic introduction/focus and topic elaboration/ clarification. Topic introduction/focus bundles give signals that a new topic is being introduced (e.g. *do you know what)*. Topic elaboration/clarification bundles focus on adding information to a topic or to relate between two units of discourse (e.g., *on the other hand*). Another function can be to ask for further clarification of previously stated information (e.g. *what do you mean*).

(3) Referential bundles make direct reference to physical or abstract entities or sometimes single out some important features of an identity to be important. There are four types of referential expressions. Identification/focus bundles identify an entity or part of it as noteworthy (e.g., *one of the most* and *that's one of the*). Imprecision bundles communicate that previous discourse is expressed imprecisely (e.g. *or something like that*). The sub-category "specification of attributes" bring focus to some particular attribute of the entity. This sub-category is further divided into three types: quantity specifications including quantities (e.g., *per cent of the*), tangible framing attributes (e.g., *in the form of*), and intangible framing attributes (e.g. *the nature of the, in the absence of* and *the way in which*). Time/place/text references can refer to one of those areas or be multi-functional (e.g. *the end of the*). Topic/subject/lecture bundles are used to make somehow direct or indirect reference to previously-mentioned issues in the lectures (e.g., *as shown in figure*). Moreover, the sub-category (D3. Text-deixis) in Biber et al.,'s framework was omitted in this study and it was replaced by a new sub-category (D3. Topic/subject/lecture reference). Multi-functional reference bundles are used to refer to a specific period of time, particular place or topic depending on context (e.g., *at*

*the end of*). The fourth category "special conversational functions" are mostly found in spoken discourse more than written discourse. This category covers three subcategories: politeness routines (e.g., *thank you very much*), simple inquiry (e.g., *what are you doing*), and reporting clauses (e.g., *I said to him*).

## 3.6 Interrater Reliability

In the present study, coding reliability analysis was carried out to assure the reliability of the coding of the lexical bundles. Therefore, a small set of corpus (20% of the corpora-10 texts per corpus) were coded by the researcher and a co-rater. The co-rater is a Master's student of English as a second language. She was given a coding framework/schema comprising explanations and examples. The researcher and the second inter-rater decided on structures and functions of the identified lexical bundles and found equivalent codes. Then, inter-rater reliability was assessed using Cohen's Kappa. The result of Cohen's Kappa was 0.82 for inter-rater reliability which showed substantial agreement.

## 3.7 Summary of the Chapter

To sum up, this chapter explains research design that includes the BASE corpus and analytical framework used for this study. It also discusses the data analysis procedures and categorization for lexical bundles used in the study. The subsequent chapter will discuss the data analysis and findings in line with the research questions.

# CHAPTER 4: RESULTS AND DISCUSSION

## 4.1 Introduction

The purpose of this chapter is to present and discuss the findings of the corpus of academic lectures of two broad disciplinary divisions namely, Arts and Humanities (AH), and Life and Medical Sciences (LS). The research questions are listed below:

1. What are the most frequent four-word lexical bundles found in the academic lectures of Arts and Humanities (AH), and Life and Medical Sciences (LS) disciplines?

2. How are the academic lectures in Arts and Humanities, and Life and Medical Sciences disciplines different or similar in terms of the structural characteristics and discourse functions of the lexical bundles used?

Research question (1) will be answered in the following section.

## 4.2 The Frequency of Lexical Bundles in the Academic Lectures of Arts and Humanities and Life and Medical Sciences

Lexical bundles in this corpus were identified by using the computer software AntConc3.4.0 that was also used to make a list of the most frequent four-word lexical bundles. In order to identify the target bundles in the lectures, certain cut-off frequencies as well as dispersion criteria were needed. Previous literature had been referred to in making the decision and within the confines of this study, the cut-off point was set at 20 times per million words with the dispersion of occurrence in at least four different lectures. These lexical bundles which met this criteria were annotated and listed for analysis.

As can be seen in Table 4.1, a list of the top 20 most frequent lexical bundles was extracted from the whole corpus. As shown in the table, the bundle *the end of the* was the most frequently-used bundles in the whole corpus occurring 96 times (15%). Bundles *at the end of* and *the way in which* accounted for the next commonly-used bundles, and occurred 88 (13.8%) and 87 times (13.6%) respectively. It was also found that seventeen out of the twenty lexical bundles in the list in Table 4.1 had also been identified by Biber et al. (2004) in his sub-corpus of classroom teaching discourse.

**Table 4.1: The 20 most frequently used lexical bundles in the whole corpus**

| | Lexical Bundles | Arts and Humanities (%) | Life and Medical Sciences (%) | Total (%) |
|---|---|---|---|---|
| **1** | the end of the | 60(9.4%) | 36(5.6%) | 96(15%) |
| **2** | at the end of | 49(7.7%) | 39(6.1%) | 88(13.8%) |
| **3** | the way in which | 77(12.1%) | 10(1.5%) | 87(13.6%) |
| **4** | to be able to | 22(3.4%) | 47(7.3%) | 69(10.8%) |
| **5** | if you look at | 37(5.8%) | 24(3.7%) | 61(9.5%) |
| **6** | if you want to | 15(2.3%) | 45(7%) | 60(9.4%) |
| **7** | going to talk about | 10(1.5%) | 48(7.5%) | 58(9.1%) |
| **8** | you can see that | 14(2.2%) | 41(6.4%) | 55(8.6%) |
| **9** | and you can see | 19(2.9%) | 35(5.5%) | 54(8.4%) |
| **10** | i'm not going to | 14(2.2%) | 40(6.2%) | 54(8.4%) |
| **11** | one of the things | 28(4.4%) | 26(4%) | 54(4.8%) |
| **12** | at the same time | 30(4.7%) | 17(2.6%) | 47(7.3%) |
| **13** | in the case of | 11(1.7%) | 36(5.6%) | 47(7.3%) |
| **14** | is going to be | 14(2.2%) | 33(5.1%) | 47(7.3%) |
| **15** | what I want to | 31(4.8%) | 15(2.3%) | 46(7.2%) |
| **16** | on the other hand | 31(4.8%) | 13(2%) | 44(6.9%) |
| **17** | and this is the | 18(2.8%) | 25(3.9%) | 43(6.7%) |
| **18** | is one of the | 30(4.7%) | 12(1.8%) | 42(6.6%) |
| **19** | are going to be | 13(2%) | 28(4.4%) | 41(6.4%) |
| **20** | in terms of the | 26(4%) | 15(2.3%) | 41(6.4%) |

A closer look at the list shows that there are two pairs of bundles that can be combined to form some of the most frequently occurring five word bundles. The first pairs of bundles are: *the end of the: at the end of* (occurring 88 times, 13.8%) + *the end of the*

(occurring 96 times, 15%) = *at the end of the*. The bundle, *the end of the,* occurred 60 times (9.4%) in AH and 36 times (5.6%) in LS, and totalling 96 times (15%) in the whole corpus. Similarly, the bundle *at the end of* was used 49 times (7.7%) in AH lectures and 39 times (6.1%) in the lectures of LS which equals 88 times (13.8%) in both corpora.

The second pair of bundles *and you can see: and you can see* (occurring 54 times, 8.4%) + *you can see that* (occurring 55 times, 8.6%) = *and you can see that.* The lexical bundle *and you can see* occurred 19 times (2.9%) in AH lectures and as frequent as 35 times (5.5%) in the lectures of LS and a total of 54 times (8.4%) in the whole corpus. Similarly, the bundle *you can see that* was used only 14 times (2.2%) in the lectures of AH and it was more frequently used in the LS lectures with a total of 41 times (6.4%) and 55 times (8.6%) in the whole corpus.

Table 4.1 shows the 20 most frequent lexical bundles in the corpus, all of which occurred at least 20 times in each broad disciplinary grouping. It is worth noting here that the lexical bundle (*per cent of the*) appeared in the list of the top 20 most frequently used lexical bundles in the lectures of Arts and Humanities, and Life and Medical Sciences. This lexical bundle occurred only in the Life and Medical Sciences lectures and it was ranked in the 12$^{th}$ place. The bundle (*per cent of the*) occurred 49 times per on million words (7.7%) and was found in 20 lectures. On the other hand, this bundle did not occur in the other sub-corpus of Arts and Humanities, therefore it was removed from the list of the top 20 most frequent lexical bundles because it was considered as a discipline-specific bundle.

In addition, the identified lexical bundles in Arts and Humanities, and Life and Medical Sciences were collapsed into one single list. Those lexical bundles that occurred in only one sub-corpus and not in the other were not included in the top 20 list. Such bundles were considered as discipline-specific lexical bundles (i.e., specific to Life and

Medical Sciences disciplinary group). Therefore the lexical bundles that occurred more than 20 times per one million words and across the two disciplinary areas were considered as general lexical bundles.

### 4.2.1 Lexical Bundles in Academic Lectures of Arts and Humanities

The Arts and Humanities sub-corpus consists of 40 university lecture transcripts across a variety of disciplines. The average length of the transcripts is between 5,524 words to 10,837 words and the total number of words is 306,652. Table 4.2 illustrates the information about this sub-corpus:

**Table 4.2: Description of Arts and Humanities sub-corpus**

| Number of transcripts | Length of transcripts | Total number of words |
|---|---|---|
| 40 | 10,837-5,524 | 306,652 |

Some bundles were excluded from the final list of the retrieved bundles as they were too speech-dependent or content-dependent. Examples of speech-dependent bundles which were excluded contained repetition of some words which were characteristics of speech e.g., *it is it is, that is that is, of the of the, on the on the,* and *in a in a*. These short phrases were called "repeats" by Biber et al. (1999) which represented a kind of hesitation phenomena that were typical of spontaneous speaking (Wang, 2017). These phrases were only used as pause fillers or hesitation markers in the lectures and could not be considered as lexical bundles (Kashiha, 2015). These short phrases were also excluded in some other studies (Nesi & Basturkmen, 2006; Kashiha & Chan, 2013, 2014; Kashiha, 2015). Similarly, bundles which were too specific to the content were also excluded such as *the First World War, the Second World War,* and *the development of capitalism*. These bundles and other similar bundles could only be used when talking about a particular topic. Appendix A gives a list of bundles which were excluded.

**Table 4.3: Frequency information on Arts and Humanities sub-corpus**

| Sub-corpus | Bundle types | Bundle tokens | % of total words |
|---|---|---|---|
| Arts and Humanities | 309 | 3221 | 0.54% |

The total number of different bundles which met the criterion set for the sub-corpus was 309 types, totalling 3221 tokens which accounted for 0.54% of the total words in the corpus as shown in Table 4.3. Appendix B provides the full list of target bundles for analysis in the Arts and Humanities sub-corpus. The occurrence of these bundles in academic lectures of AH suggests that they rely heavily on using different lexical bundle types and this may be explained because of the possibility that the explanatory nature of this disciplines requires that the ideas must be connected especially in argumentations. This requires the lecturers to use a variety of multi-word chunks in order to convey their messages.

The bundle *the way in which* was by far the most frequent lexical bundles in the sub-corpus, occurring 77 times per one million words (24.9%) and was found in 20 lectures 50% of the total number of lectures. The second most frequent bundle, *the end of the* occurred 60 times (19.4%) and in 29 different lectures (72%) which showed that this lexical bundle was wider distributed across lectures. Table 4.4 shows the list of the top 10 most frequent lexical bundles in the Arts and Humanities lectures.

**Table 4.4: The top ten most frequent 4-word lexical bundles in Arts and Humanities**

| No. | Lexical bundles | Frequency (%) | No. of lectures (%) |
|---|---|---|---|
| **1** | the way in which | 77(24.9%) | 20(50%) |
| **2** | the end of the | 60(19.4%) | 29(72%) |
| **3** | at the end of | 49(15.8%) | 25(62.5%) |
| **4** | if you look at | 37(11.9%) | 20(50%) |
| **5** | on the other hand | 31(10%) | 18(45%) |
| **6** | what i want to | 31(10%) | 21(52.5%) |
| **7** | at the same time | 30(9.7%) | 17(42.5%) |
| **8** | in other words the | 30(9.7%) | 18(45%) |
| **9** | is one of the | 30(9.7%) | 17(42.5%) |
| **10** | on the one hand | 29(9.3%) | 13(32.5%) |

### 4.2.2 Lexical Bundles in Academic Lectures of Life and Medical Sciences

The sub-corpus Life and Medical Sciences consists of 40 university lecture transcripts with the average length of the transcripts between 3,044 words to 14,978 words and the total number of words is 288,226. Table 4.5 illustrates the information about this sub-corpus:

**Table 4.5: Description of Life and Medical Sciences sub-corpus**

| Number of transcripts | Length of transcripts | Total number of words |
|---|---|---|
| 40 | 14,978-3,044 | 288,226 |

From the list of retrieved bundles, some bundles were omitted from the final list of the retrieved bundles as they were too speech-dependent or content-dependent. Bundles such as *in the in the* and *if you if you* were excluded as these were considered examples of speech-dependent bundles which contained repetition of some words. These bundles were only used as pause fillers in the lectures and could not be considered as lexical bundles. Some other bundles were also excluded such as defining quantities: *fifty per cent of, twenty per cent of, ten per cent of,* and *five per cent of*, and some other bundles being too specific to the content: *protein in the urine, the clinical methods course,* and *the*

*development of the*. Appendix C presents a list of bundles which were excluded from the final list of bundles in Life and Medical Sciences.

**Table 4.6: Frequency information on Life and Medical Sciences sub-corpus**

| Sub-corpus | Bundle types | Bundle tokens | % of total words |
|---|---|---|---|
| Life and Medical Sciences | 327 | 3388 | 0.56% |

After the identified expressions were removed, altogether, there were 3388 individual bundles in the sub-corpus including 327 different bundle types, as can be seen in Table 4.6. The high occurrence of these bundles in academic lectures suggests that the lecturers in LS sub-corpus were highly dependent on demonstrating and reporting experiments/ calculations/ measurements. Appendix D gives the full list of target bundles for analysis in Life and Medical Sciences.

Results of the analysis show that the most frequent lexical bundle found in the lectures of LS is the bundle *going to talk about*. Table 4.7 presents the list of the 10 most frequent bundles in the sub-corpus. As can be seen, the bundle *going to talk about* has a frequency of 48 raw occurrences (14.6%) and occurs in 22 lectures (55%). The second most frequent bundle *to be able to* is used in 21 lectures (52.5%) with a raw frequency of 47 times (14.3%). As can be seen, the bundle *if you want to* is the third most frequently-used bundle in the LS sub-corpus occurring 45 times (13.7%). The list of bundles in Table 4.7 also shows that there are two pairs of bundles that make up the most frequent 5-word bundle. The first pair of bundles is*: at the end of* (occurring 39 times 11.9%) + *the end of the (*occurring 36 times 11%) = *at the end of the*. The second pair of bundles: *and you can see* (occurring 35 times 10.7%) + *you can see that* (occurring 41 times 12.5%) = *and you can see that.*

**Table 4.7: The top ten most frequent 4-word lexical bundles in Life and Medical Sciences**

| No. | Lexical bundles | Frequency (%) | No. of lectures (%) |
|---|---|---|---|
| **1** | going to talk about | 48(14.6%) | 22(55%) |
| **2** | to be able to | 47(14.3%) | 21(52.5%) |
| **3** | if you want to | 45(13.7%) | 22(55%) |
| **4** | per cent of the | 45(13.7%) | 20(50%) |
| **5** | you can see that | 41(12.5%) | 15(37.5%) |
| **6** | i'm not going to | 40(12.2%) | 21(52.5%) |
| **7** | at the end of | 39(11.9%) | 17(42.5%) |
| **8** | in the case of | 36(11%) | 07(17.5%) |
| **9** | the end of the | 36(11%) | 18(45%) |
| **10** | and you can see | 35(10.7%) | 16(40%) |

### 4.2.3 Comparison between the Frequencies of the Lexical Bundles in the Corpora

A comparison of frequency of occurrence of the bundles found in this study reveals some marked differences of lexical bundles in both corpora. The most ostensible differences are the number of bundle types and tokens in each disciplinary grouping used in the corpora. Table 4.8 gives the detailed information of lexical bundles. As can be seen, out of 6609 individual cases of four-word lexical bundles which meets the identification criteria, 3221 tokens belong to the sub-corpus of AH, whereas, LS lectures uses a slightly higher number which accounts for 3388 tokens of the bundles. A similar pattern of use is reported regarding the number of bundles types. AH lectures use 309 different bundle types and LS lectures use 237 of bundle types.

**Table 4.8: Frequency information of lexical bundles in the corpora**

| Corpus | Bundle types | Bundle token | % of total words |
|---|---|---|---|
| Arts and Humanities (AH) | 309 | 3221 | 0.54% |
| Life and Medical Sciences (LS) | 327 | 3388 | 0.56% |
| **Total** | 636 | 6609 | 1.1% |

Another difference between the corpora is the type of bundles that have been removed from each sub-corpus. In AH lectures, most of the excluded bundles are those speech dependent with words or letters repeating currently as gap fillers or markers of hesitation in speech *(*e.g.*, it is it is* and *that is that is)*. The number of discipline-dependent nouns of specific topics or subjects related to the disciplines are also found in these lectures *(*e.g., *the Second World War, the development of capitalism).* In AH disciplines, the speech dependent bundles occurr more frequently than content-dependent bundles.

In LS lectures, speech dependent bundles are excluded more than the content-dependent bundles. Also most of the excluded content dependent bundles are those containing defining quantities *(*e.g., *fifty per cent of)* and discipline-dependent content *(*e.g., *the clinical methods course)*. In addition, the topics that are discussed in LS lectures are experimental which would need the use of numbers and certain quantifiers as well as specific acronyms and jargons.

In addition, the removed bundles in both corpora are mostly speech-dependent; they are considered speech fillers and are not considered for analysis. This suggests that both AH lectures and LS lectures rely on gap fillers and other types of speech markers in delivering lectures. Conversely, AH lectures contain more content and speech-dependent bundles than LS lectures.

In addition, another difference is the topics which have been discussed in each sub-corpus. The higher occurrence of lexical bundle tokens in LS is due to the fact that the subjects belonging to LS such as biology and agricultural botany may use more lexical bundles that give cohesion to the ideas when dealing with the description, prediction, and understanding of natural phenomena, based on empirical evidence from observation and experimentation. On the other hand, AH sub-corpus is more inductive and subjects like history and philosophy may use more elaborations to present views, and this requires the

lecturer to use multi-word combinations in order to give cohesion to those ideas and make the topics discussed as comprehensive as possible for the target audience.

Lectures in LS contained a large number of types and tokens of bundles than lectures in AH. The higher occurrence of lexical bundles in the LS sub-corpus could be taken to mean that lectures in the science show a higher preference for the use of ready-made lexical sequences (e.g., *i'm going to talk, and things like that* and *you can see here*) to deliver the message. Using these lexical bundles also appear to lead to word economy as these lectures are shown to have the lowest number of words, as LS disciplines are largely dependent on numbers or images in presenting data (e.g., *if we look at, it's one of the* and *per cent of people*).

LS disciplines share some features with disciplines that belong to Physical Sciences, in that, some disciplines such as biological sciences may belong to both of these disciplinary divisions either to Life and Medical sciences or Physical Sciences. Therefore disciplines in LS may have some features that are similar to those found in disciplines of physical sciences such as the use of visual aids and images to support their findings. The LS disciplines are more explanatory, less vigorous and practical (Hyland, 2008).

These findings are in line with Hyland's (2008) conclusion that applied and pure sciences *(*e.g., *electrical engineering, Microbiology)* show a greater preference for the use of prefabricated expressions than disciplines in the social sciences *(*e.g., *business studies, applied linguistics)* and he reasons that "it could be a consequences of the relatively abstract and graphical nature of technical communication (p. 9). This is evident in the disciplines like Mathematics and Statistics which are more technical disciplines. These disciplines are largely dependent on numbers or images in presenting data and more formulaic ways in the use of words (e.g., *per cent of the, you can see that, so if you look* and *when we look at*).

Disciplines in LS *(*e.g., *medicine, biological sciences, zoology, and animal and microbial sciences,* and *agricultural botany)* deal mainly with human beings, animals, plants and other natural phenomenon. However, other disciplines *(mathematics and statistics)* have a more physical nature in hard sciences, and deal mainly with objects and numbers rather than people. Based on this fact, these lectures are more likely to give explicit results in a more direct way. One way to achieve cohesion is through using prefabricated chunks which students are already familiar with.

### 4.2.4 Patterns and Variations

There have been remarkable variations in the use of four-word lexical bundles. Out of the 636 different bundle types identified in the whole corpus, there are 105 shared bundles between the sub-corpora (Appendix E) (e.g., *the way in which, the end of the, on the other hand, if you look at* and *what we're going to)*.

However, lecturers in LS do not tend to use 204 of the target bundles which seem to be exclusively used by the AH sub-corpus (e.g., *on the one hand, the idea of the, all the rest of, as a kind of,* and *so in other words)* (see Appendix F). On the other hand, LS lectures use 218 lexical bundles which are different from those found in AH lectures (e.g., *I'm going to talk, and things like that, per cent of people, quite a lot of,* and *so you have to*) (see Appendix G). In other words, academic discourse is characterized by the use of these word combinations in a range of academic registers or disciplines such as *on the other hand* to show topic shift or *if you look at* to raise awareness. The other identified bundles are either shared between the two disciplinary groups of lectures or specific to either of the groups.

Despite some disciplinary differences, these large number of common bundle types in the two sub-corpora emphasize the high reliance of university lectures as

representative of academic genre on multi-word combinations in general and lexical bundles in specific to convey their messages. The most frequent 4-word lexical bundles in the two sub-corpora, with the shared bundles (in bold) are listed in Table 4.9. As can be seen, only 19 of the top 50 lexical bundles in each sub-corpora occurr in the list of the other, which is indicative of the inclination that lecturers in different disciplinary fields rely on different resources to convince their audiences and communicate their disciplinary messages.

**Table 4.9: Most frequent 50 four-word lexical bundles in both corpora**

| | Arts and Humanities | Frequency (%) | Life and Medical Sciences | Frequency (%) |
|---|---|---|---|---|
| **1** | the way in which | 77(24.9%) | going to talk about | 48(14.6%) |
| **2** | **the end of the** | 60(19.4%) | **to be able to** | 47(14.3%) |
| **3** | **at the end of** | 49(15.8%) | **if you want to** | 45(13.7%) |
| **4** | **if you look at** | 37(11.9%) | per cent of the | 45(13.7%) |
| **5** | on the other hand | 31(10%) | you can see that | 41(12.5%) |
| **6** | **what i want to** | 31(10%) | **i'm not going to** | 40(12.2%) |
| **7** | **at the same time** | 30(9.7%) | **at the end of** | 39(11.9%) |
| **8** | in other words the | 30(9.7%) | in the case of | 36(11%) |
| **9** | is one of the | 30(9.7%) | **the end of the** | 36(11%) |
| **10** | on the one hand | 29(9.3%) | **and you can see** | 35(10.7%) |
| **11** | and so on and | 28(9%) | **is going to be** | 33(10%) |
| **12** | **one of the things** | 28(9%) | i'm going to talk | 30(9.1%) |
| **13** | **in terms of the** | 26(8.4%) | a little bit of | 28(8.5%) |
| **14** | and one of the | 25(8%) | are going to be | 28(8.5%) |
| **15** | the idea of the | 24(7.7%) | if we look at | 26(7.9%) |
| **16** | all the rest of | 23(7.4%) | **one of the things** | 26(7.9%) |
| **17** | as a kind of | 23(7.4%) | **and this is the** | 25(7.6%) |
| **18** | the rest of it | 23(7.4%) | going to look at | 25(7.6%) |
| **19** | **to be able to** | 22(7.1%) | **if you look at** | 24(7.3%) |
| **20** | and all the rest | 21(6.7%) | and things like that | 23(7%) |
| **21** | at the beginning of | 21(6.7%) | of the things that | 23(7%) |
| **22** | the rest of the | 21(6.7%) | **and i'm going to** | 22(6.7%) |
| **23** | **and you can see** | 19(6.1%) | a little bit about | 21(6.4%) |
| **24** | so in other words | 19(6.1%) | it's going to be | 20(6.1%) |
| **25** | to do with the | 19(6.1%) | thank you very much | 20(6.1%) |
| **26** | **a lot of the** | 18(5.8%) | **and this is a** | 19(5.8%) |
| **27** | **and this is the** | 18(5.8%) | **i don't know if** | 19(5.8%) |
| **28** | **i don't know if** | 18(5.8%) | per cent of people | 19(5.8%) |
| **29** | in the context of | 18(5.8%) | **what i'm going to** | 19(5.8%) |

| | | | | |
|---|---|---|---|---|
| **30** | **you look at the** | 18(5.8%) | we're going to look | 18(5.5%) |
| **31** | i want to do | 17(5.5%) | **at the same time** | 17(5.1%) |
| **32** | if you like the | 17(5.5%) | i just want to | 17(5.1%) |
| **33** | in the way that | 17(5.5%) | if you have a | 16(4.8%) |
| **34** | so on and so | 17(5.5%) | quite a lot of | 16(4.8%) |
| **35** | **and i'm going to** | 16(5.1%) | so i'm going to | 16(4.8%) |
| **36** | as a sort of | 16(5.1%) | so this is a | 16(4.8%) |
| **37** | is a kind of | 16(5.1%) | **a lot of the** | 15(4.5%) |
| **38** | **what i'm going to** | 16(5.1%) | i want you to | 15(4.5%) |
| **39** | you know this is | 16(5.1%) | **in terms of the** | 15(4.5%) |
| **40** | if you like of | 15(4.8%) | so you have to | 15(4.5%) |
| **41** | **if you want to** | 15(4.8%) | there's a lot of | 15(4.5%) |
| **42** | on and so forth | 15(4.8%) | those of you who | 15(4.5%) |
| **43** | one of the most | 15(4.8%) | what do you think | 15(4.5%) |
| **44** | the nature of the | 15(4.8%) | **what i want to** | 15(4.5%) |
| **45** | **and this is a** | 14(4.5%) | you can see here | 15(4.5%) |
| **46** | at the heart of | 14(4.5%) | **you look at the** | 15(4.5%) |
| **47** | **i'm not going to** | 14(4.5%) | you need to know | 15(4.5%) |
| **48** | **is going to be** | 14(4.5%) | you're going to be | 15(4.5%) |
| **49** | one of the reasons | 14(4.5%) | can see that the | 14(4.2%) |
| **50** | the beginning of the | 14(4.5%) | going to be a | 14(4.2%) |

The possible explanation for such a large number of shared bundles may be rooted in the fact that in academia, a common set of word combinations are commonly used which can be regarded as an indispensable part and thus largely appear in the language use of different disciplinary communities. Moreover, shared bundles such as: *if we look at, and you can see,* and *in terms of the* may emerge as similar characteristics of most of the disciplines.

On the other hand, some marked variations include that AH disciplines are representative of soft fields in which the characteristics are quite different from LS disciplines that are grouped under applied and pure sciences. This contrast makes some marked variations in language use of the two groups of academic domains with those in AH providing information and ideas in the form of definition and explanation (e.g., *in other words, the idea of the* and *the nature of the),* and LS requiring to use some word

strings to point to graphs or tables and other visual aids and tools in the teaching (e.g., *you can see here, you look at the*, and *we're going to look)*.

Besides the found similarities, each sub-corpus is characterized by a number of specific bundles. LS lectures encompass the larger number of exclusive bundles with 218 types. This is followed by AH with 204 types (see Appendix F &G). A possible justification for such a diversity of use may be the varied nature of each science which requires lecturers to use particular bundles relating to the topic under discussion.

One way to realize the variations in the language use of different disciplinary divisions is their selection of lexical bundles. The fact that lectures in each disciplinary divisions rely on a specific set of word combinations (which were not found in the other sub-corpus) would support the idea that they have their own peculiarities and ways of organizing the discourse to get their disciplinary messages across. Table 4.10 lists the top 5 bundles specific to the sub-corpus of each division.

**Table 4.10: The top 5 most frequent 4-word lexical bundles specific to each corpus**

| | **Arts and Humanities** | **Frequency** |
|---|---|---|
| **1** | On the other hand | 31(10%) |
| **2** | the idea of the | 24(7.7%) |
| **3** | all the rest of | 23(7.4%) |
| **4** | as a kind of | 23(7.4%) |
| **5** | I want to do | 17(5.5%) |
| | **Life and Medical Sciences** | **Frequency** |
| **1** | per cent of the | 45(13.7%) |
| **2** | i'm going to talk | 30(9.1%) |
| **3** | and things like that | 23(7%) |
| **4** | per cent of people | 19(5.8%) |
| **5** | we're going to look | 18(5.5%) |

In this regard, a few of the bundles which are specific to AH are used to fulfill one of the communicative purposes of the disciplines presented in AH. Take history lectures for

example, which apply a number of 4-word bundles such as: *the second half of, towards the end of*, and *for a long time* to refer to specific time period in the past in their discussion. This feature seems to be characteristics of the discipline of history to use a large number of time markers compared to other disciplines. Many other bundles are found to occur only in the corpus of LS, dealing with people's health and medicine. Some discipline-specific bundles used to quantify and qualify this relationship are *per cent of people, per cent of all*, and *one of the problems.* The rest of the specific bundles in LS are generally used to convey some other communicative purposes such as *we're going to do, to give you some* and *there may be a*.

In each sub-corpus, most of the specific bundles are used generally to convey a communicative function related to a discipline and a topic in concern. However, a closer look at the list of specific bundles in each sub-corpus shows some noticeable disciplinary variations in the use of particular words (Table 4.10). An interesting phenomenon is that the most frequent 4-word lexical bundles in AH belong to the structural category called noun phrase with of-phrase fragment. Another important disciplinary variation is the use of first person singular/ plural pronouns *(I, we)* and second person singular pronoun (you) in both AH and LS lectures. LS lectures use slightly more personal pronouns than the AH lectures (Table 4.11).

## 4.3 Structural Analysis of Lexical Bundles in Academic Lectures

This section answers the first part of the second research question:

2. How are the academic lectures in Arts and Humanities, and Life and Medical Sciences disciplines different or similar in terms of the structural characteristics and discourse functions of the lexical bundles used?

The identified four-word lexical bundles are categorized structurally using the classification proposed by Biber, et al. (2004). Their structural framework includes three main structural types for lexical bundles: 1) lexical bundles that incorporate verb phrase fragments such as *is based on the*; 2) lexical bundles that incorporate dependent clause fragments like *I want you to*; and 3) lexical bundles that incorporate noun phrase and prepositional phrase fragments such as *at the end of* (Biber et al., 2004, p.381). Each main grammatical type has a number of sub-structures which will be listed and discussed below.

A first glance at the results indicates that lexical bundles identified in the corpora of the two broad disciplinary divisions appear to have a number of similarities and differences in the structural types of the bundles used by the lecturers. The following sections present the frequency distribution of the lexical bundles found in the corpus of each disciplinary division followed by a comparison between the two corpora in terms of the lexical bundles used. These analyses answer the research question 2.

### 4.3.1 Structural Distribution of Lexical Bundles in Arts and Humanities Lectures

Table 4.11 shows the distribution of lexical bundles across the structural types in AH lectures. This table demonstrates the detailed percentages of the three main structural categories with their specific sub-categories in the lectures of both corpora using the classification proposed by Biber et al. (2004, p. 381). A primary examination of the corpora indicates that AH lectures used all the three main grammatical types, which show the lecturers' flexibility in selecting different structures in constructing lexical bundles. The main purpose of this variation of use is to better communicate the content of the lectures and thus facilitate the students' comprehension of specific disciplinary messages.

**Table 4.11: Distribution of lexical bundles across structural types in Arts and Humanities**

| Structure categories | Sub-categories | No. % of bundles |
|---|---|---|
| 1. verb-phrase fragments | 1a. (connector+) 1st+2nd person+ VP fragments | 37(11.9%) |
| | 1b. (connector+) 3rd person pronoun + VP fragment | 30(9.7%) |
| | 1c. discourse marker + VP fragment | 8(2.5%) |
| | 1d. Verb phrase (with non-passive verb) | 28(9.0%) |
| | 1e. Verb phrase (with passive verb) | 1(0.3%) |
| | 1f. yes-no question fragments | 2(0.6%) |
| | 1g. WH- question fragments | 1(0.3%) |
| **Sub-total** | | **107(34.3%)** |
| 2. dependent clause fragments | 2a. 1st/2nd person pronoun + dependent clause fragment | 9(2.9%) |
| | 2b. WH-clause fragments | 8(2.5%) |
| | 2c. If-clause fragments | 11(3.5%) |
| | 2d. (verb+adjective) to-clause fragment | 19(6.1%) |
| | 2e. That-clause fragment | 7(2.2%) |
| **Sub-total** | | **54(14.3%)** |
| 3. noun phrase +prepositional fragments | 3a. (connectors+) Noun phrase with of-phrase fragment | 58(18.7%) |
| | 3b. Noun phrase with other post-modifier fragment | 11(3.5%) |
| | 3c. Other noun phrase expressions | 17(5.5%) |
| | 3d. Prepositional phrase expressions | 59(19.0%) |
| | 3e. Comparative expressions | 3(0.9%) |
| **Sub-total** | | **148(47.6%)** |
| **Total** | | **309(98.9%)** |

Analysis of the data indicates that lexical bundles in the sub-corpus of AH makes use of all three major structural categories in the taxonomy. Table 4.11 gives the detailed information about the structural distribution of the bundles in the AH sub-corpus. As it is shown in the table, noun and prepositional phrase fragments (e.g., *the nature of the, on the other hand* and *in terms of the)* act as a leading category in the sub-corpus. These

structures account for almost half of the bundles (47.6%) in the AH sub-corpus. The structural category that comes in the second place with 34.3% is the verb phrase fragments (e.g., *is going to be, take a look at* and *was a kind of*) and dependent clause fragments (e.g., *I don't know if, if we look at* and *want to do is)* with 14.3%.

As for the sub-category (noun phrase + prepositional fragments) results indicate that prepositional phrase expressions *(*e.g., *at the level of, of the things that* and *at the top of)* and noun phrase with of-phrase fragments *(*e.g., *the end of the, one of the most* and *a little bit of)* are the most frequently-used lexical bundles which comprised structures of the noun phrase category with 19% and 18.7% respectively. The sub-category 1st + 2nd person+ verb-phrase fragments is the most frequently used (with 11.9% of lexical bundle types).This examples include: *I'm not going to, you know if you* and *we're going to be.* This is followed by the 3rd person pronoun + verb-phrase fragments (with 9.7%), for example: *and this is a, this is one of the,* and *it has to be*. The less frequent category is the (dependent clause fragments). Results show that the most frequent structural sub-category is the verb/adjective + to-clause fragments with a percentage occurrence of 6.1%, with examples including *to come up with, want to do is* and *to be able to*.

### 4.3.2 Structural Distribution of Lexical Bundles in Life and Medical Sciences

All lexical bundles in LS sub-corpus have been classified according to the structural taxonomy suggested by Biber et al., (2004). Table 4.12 presents the percentage of these bundles in each sub-category.

**Table 4.12: Distribution of lexical bundles across structural types in Life and Medical Sciences**

| Structure categories | Sub-categories | No. % of bundles |
|---|---|---|
| 1. verb-phrase fragments | 1a. (connector+) 1st+2nd person+ VP fragments | 66(20.1%) |
| | 1b. (connector+) 3rd person pronoun + VP fragment | 32(9.7%) |
| | 1c. discourse marker + VP fragment | 1(0.3%) |
| | 1d. Verb phrase (with non-passive verb) | 45(13.7%) |
| | 1e. Verb phrase (with passive verb) | ----- |
| | 1f. yes-no question fragments | 5(1.5%) |
| | 1g. WH- question fragments | 3(0.9%) |
| **Sub-total** | | **152(46.2%)** |
| 2. dependent clause fragments | 2a. 1st/2nd person pronoun + dependent clause fragment | 15(4.5%) |
| | 2b. WH-clause fragments | 12(3.6%) |
| | 2c. If-clause fragments | 23(7.0%) |
| | 2d. (verb+adjective) to-clause fragment | 21(6.4%) |
| | 2e. That-clause fragment | 13(3.9%) |
| **Sub-total** | | **84(25.4%)** |
| 3. noun phrase +prepositional fragments | 3a. (connectors+) Noun phrase with of-phrase fragment | 38(11.6%) |
| | 3b. Noun phrase with other post-modifier fragment | 12(3.6%) |
| | 3c. Other noun phrase expressions | 11(3.3%) |
| | 3d. Prepositional phrase expressions | 29(8.8%) |
| | 3e. Comparative expressions | 1(0.3%) |
| **Sub-total** | | **91(27.6%)** |
| **Total** | | **327(99.2%)** |

Analysis of the data shows that lexical bundles in LS lectures make use of all three major structural categories in the taxonomy. Table 4.12 presents the detailed information about the percentage of structural distribution of the bundles in the LS sub-corpus. As indicated in Table 4.12, verb phrase fragments *(*e.g., *are going to be, and I'm going to* and *going to look at)* act as a leading category in the corpus (46.2%) which accounts for almost half of all lexical bundles types that occur in this sub-corpus. The rest of the

bundles are shared almost equally between the two other structure categories with 27.6% in the noun phrase + prepositional fragments *(*e.g., *the way in which, the end of the* and *and things like that)* and 25.4% dependent clause fragments *(*e.g., *if you look at, what I want to* and *that there is a).*

As for the sub-categories, results indicate that the sub-category:1st + 2nd person+ verb-phrase fragments are the most frequently-used (with 20.1% of lexical bundle types), with examples including: *I'm not going to, you know if you* and *we're going to be,* and verb phrase (with non-passive verbs), as in: *go back to the, not going to go* and *is one of the* are the most comprised structures of the verb phrase fragments category with 13.7%. The sub-category noun phrase with of-phrase fragments is the most frequently used (with 11.6% of lexical bundle types) with examples that include *and one of the, a lot of people* and *this part of the,* and prepositional phrase expressions with 8.8%, with the following examples: *on the other hand, in the middle of,* and *of you who are* in the second place in the main category of noun phrase + prepositional phrase fragments. The less frequent category is the dependent clause fragments, where results show that the most frequent structural sub-category is the if-clause fragments with a percentage occurrence of 7%, such as: *if you've got a, if you're going to* and *if you do i*.

### 4.3.3 The Comparison between Structures of Lexical Bundles in the Corpora

Table 4.13 shows the overall distribution of lexical bundles in the two domains. The total number of all lexical bundles is 363. The academic lectures in AH sub-corpus account for 309 bundles, while the lectures in LS contain a considerably higher number of lexical bundles with 327 bundle types. The lectures in AH use a slightly lower number of lexical bundle types than the lectures in LS, suggesting that the language in the LS lecturer is more varied.

**Table 4.13: Distribution of lexical bundles across the main structural types in the corpora**

| Structural categories | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| Verb-phrase fragment | 107(34.3%) | 152(59.9%) |
| Dependent clause | 54(14.3%) | 84(25.4%) |
| NP and PP fragments | 148(47.6%) | 91(27.6%) |
| **Total** | **309(98.9%)** | **327(99.2%)** |

The distribution of bundles across the structural categories of lexical bundles in the two domains is illustrated in Table 4.13. It turns out that the distributions of lexical bundles across the structural categories in AH and LS do not differ dramatically. The most frequent category in AH sub-corpus is the noun phrase+ prepositional phrase fragment, while in LS sub-corpus it is the verb-phrase fragment. The second most dominating category in AH is the verb-phrase fragment, and the noun phrase+ prepositional phrase fragment in LS. Dependent clause takes the last place in AH and it also appears to be the least frequent in the academic lectures in LS.

From a large number of bundles in the LS sub-corpus, as is inferred from the figures in Table 4.13, lecturers in LS appear to highly rely on the use of verb phrase fragments to convey such functions and to direct the audience towards the tools (e.g., *as you can see* and *have a look at)*. Lecturers in the two groups choose verb-phrase fragments as the preferred simple and straightforward structures to best convey their disciplinary lessons. In fact lectures in LS and AH share an almost similar rate of use of dependent clause fragments such as: *if you look at* and *I want you to* though lectures in LS appear to be a bit more reliant upon the use of dependent clauses. This could be because of the fact that the lectures in LS apply visual aids such as slides and tables, and this requires the lecturers to attract the students' attention towards the tools while teaching. Lecturers in AH make use of noun phrase + prepositional phrase fragments *(*e.g., *one of the things* and *at the beginning of)* more than LS because these might have resulted from the expressive nature

of the disciplines in LS that require the lecturers to use a greater variety of short phrases, including different nouns and prepositions, to best convey their disciplinary messages. The fluctuation also could signal the typical ways that lectures in both divisions tend to orientate their audience to the intended messages and organize their information. Such diversity in the method of conveying information imposes variations in the selection of lexical bundles in the two corpora.

### 4.3.4 Structural Classifications of Lexical Bundles

As noted above, each principle structure contains a number of sub-categories which have more specific goals. In the following sub-sections, comparisons between the two corpora in terms of the sub-categories of each main structure are presented and discussed with examples extracted from the corpus to convey such functions.

#### 4.3.4.1 The Comparison of Verb-phrase Fragments in the Corpora

The two groups of lectures show some similarities and a number of variations in the employment of different types of verb phrases to deliver their disciplinary materials. Table 4.14 provides the details of the proportion of verb phrase sub-categories across the corpora. As is shown, 1st/2nd person pronoun + VP fragments category is the largest sub-category of verb phrases in the corpora. Based on these results, it seems the academic lecturers, in general, prefer to use verb phrases preceded by personal pronouns in order to form lexical bundles. This suggests the importance of using personal pronouns, especially 1st and 2nd *(I, you, we)* at the initial position of verb-phrases in academic lectures. This way, the speaker is able to better show the direction and flow of the lecture in a more controlled way through engaging himself and the students in the process of learning. Inferred from the table, such employment is more prominent in LS with 20.1% of occurrence compared to AH with 11.9%. A possible reason to justify this phenomenon

would be the more experimental nature of the LS discipline which calls for more presence of the lecturer or the involvement of the students during the course. Some examples that are found in LS lectures include: *I'm going to talk, so you have to, you need to know, but you can see, you don't need to, you think about it, you're not going to, and we're going to, we're going to talk, you can see is* and *you have to have* which are not found in the corpora of AH. The following sentences presents the concordance lines in which some of these lexical bundles appear in the LS corpus as follows:

(1) just imagine that instead of giving you numbers I was giving you a series of instructions okay how much could you handle all in one go so anyway ***i'm going to give*** you a number and then you'll see what *(LS-Plant Sciences)*

(2) and of course the problem with these insecticides is that they can accumulate in the food chain now ***you don't need to*** copy this down because you've had it on the handout that I gave you for the practical *(LS-Zoology)*

(3) storage proteins were related to this is confirmed by real measurement in space and the two other bits that haven't mentioned are if ***you can see the*** green blobs in the middle here that is our metal *(LS-Agricultural Botany)*

(4) we have our piece of D-N-A okay imagine this is the five-prime of it and this is the three-prime of it ***and we're going to*** make er some D-N-A here we're at the origin we start with a piece R-N-A *(LS-Animal and Microbial Sciences)*

**Table 4.14: Proportional distribution of verb-phrase fragments across the corpora**

| Sub-categories | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| 1a. (connector+) 1st+2nd person+ VP fragments | 37(11.9%) | 66(20.1%) |
| 1b. (connector+) 3rd person pronoun + VP fragment | 30(9.7%) | 32(9.7%) |
| 1c. discourse marker + VP fragment | 8(2.5%) | 1(0.3%) |
| 1d. Verb phrase (with non-passive verb) | 28(9.0%) | 45(13.7%) |
| 1e. Verb phrase (with passive verb) | 1(0.3%) | ---- |
| 1f. yes-no question fragments | 2(0.6%) | 5(1.5%) |
| 1g. WH- question fragments | 1(0.3%) | 3(0.9%) |
| **Sub-total** | **107(34.3%)** | **152(59.9%)** |

However, LS lecturers report a higher percentage of use of verb-phrase (with active verbs) than those in AH lecturers Table 4.14. This variance could be due to the fact that the nature of the disciplines in LS required the lecturers to use more verbal phrases to guide listeners towards the instruction. LS lecturers' preferred simple and straightforward structures to best convey their disciplinary lessons and aid comprehension of the given lectures, as in the examples: *talk a little bit, are a lot of, bear that in mind, is that if you* and *not going to talk.* Examples from the corpus are as follows:

(5) paying for training what you're trying to do is develop a business primarily as well as to develop the individual ***and you have to*** work out some sort of balance here because actually a lot of training doesn't *(LS-Plant Sciences)*

(6) she represents at the time is the tenor that's really going on right now here are some other words coming up type and antitype now ***you have to remember*** that we're studying works written by people *(AH-English)*

(7) Starling equation you can see the various factors that will cause this to be too high ***so i'm going to*** give you a list of factors that are of clinical importance that can cause oedema and the first one is the one *(LS-Animal and Microbial Sciences)*

Lecturers in AH and LS show a similar use of the structural sub-category 3rd person + verb phrase fragments, which also have a higher occurrence in LS lectures than in AH lectures. The use of these phrases in AH may be due to the explanatory nature of the disciplines in this division in which a variety of ideas and notions were presented. It seems that lecturers' needed to refer to these ideas and notions during the given lecture using bundles with 3rd person such as *(it, this)*. Other examples of bundles include: *it's one of the, that's one of the, so this is the* and *it's not just the* to provide more information about the topic being discussed, as in:

(8) well got this flight console here maybe I can do something and see what happens if I'm bold if i stand in this one place then ***it's going to be*** all right but as you already as you know I'm could actually even just *(AH-Philosophy)*

The disciplines in the LS lectures made use of these bundles because LS has more to do with intangible items such animal and human health. Such characteristics would have led the lecturers to practice more of bundles in the form of phrases beginning with 3rd person pronoun *(it, this),* as in:

(9) ***it's one of the*** commonest chronic diseases but ***it's one of the*** diseases we know least about and ***it's one of the*** diseases that seems simple but perhaps this X-ray is really a primary example why you think it might *(LS-Medicine)*

From Table 4.14, there was only one example of discourse marker + VP fragment in the corpora of LS: *I think it's the,* while there were eight examples that occurred in the AH lectures: *I think you now* and *you know this is,* as in the following examples:

(10) the creature to have eggs that look like seeds or little snails and at that time and till two years ago ***I think it's the*** case that people genuinely didn't know they didn't know why stick insect eggs took so only *(LS-Mathematics)*

(11) what there was to know this infinite ocean and he was just picking up the odd stone the odd pebble from the beach ***you know this is*** Isaac Newton you know who was a great figure of the eighteen century *(AH-English)*

Regarding the verb phrase (with non-passive verb), both AH and LS used this sub-category in a similar way. LS lecturers made use of more of these bundles than AH sub-corpus (9% and 13.7% respectively) as shown in the following examples:

(12) it's a ocean of people in front of it right so if we're going to think about the effect of all this in terms of Rome and the effect of what the intention ***is one of the*** things we talked about very beginning today *(AH-Classics)*

(13) now what we're going to find is that this will reduce let's ***have a look at*** it on the a few little spots of blue will appear oh I need something to measure out one ml excuse me now when I put one ml of this *(LS-Mathematics)*

Concerning the yes-no questions fragments, such use can be taken to mean that the lectures exercised more disposition towards providing feedback from the students. This might have possibly resulted from the more constantly making sure whether students follow the discussion of the lectures. Eliciting information by raising such questions could help them frame the information in the lecture in a more organized way. Such yes-no question attracted audience attention towards the topic, as in the following extracts from the corpus:

(14) if you exercise your skin will feel tight over the muscles compared to normal and that's due to this tissue swelling ***are there any questions*** about this so far okay well these are two cases where thing *(LS-Animal and Microbial Sciences)*

(15) the research in problem based learning and say it supports another style of learning where actually it's something completely different ***does that make sense*** and is that enough on that is that what you were *(LS-Medicine)*

And WH-questions fragments in LS lectures as in the following bundle:

(16) and this is one of the great uncertainties really is that we not really sure ***what's going to happen*** to rainfall in the coming centuries and things like maize et cetera are very sensitive to rainfall plus or minus ten *(LS-Plant Sciences)*

(17) ***what do you think*** about the sociology modules which are which are er taught very early on in the course and that is usually a good discussion point realistic er commitment to a career in medicine *(LS-Medicine)*

(18) these films either explicitly as in Persian or implicitly in many of them are actually about ***what does it mean*** to be an Italian and that was of course a question of particular sensitivity when these films *(AH-Film and Television)*

#### 4.3.4.2 The Comparison of Dependent Clause in the Corpora

The second main category is dependent clause fragments that were higher in LS as shown in Table 4.15. LS lecturers used higher percentages of dependent clause fragments, with 25.4% compared with 14.3% in AH, suggesting the practical notion of LS disciplines which entailed more demonstration or illustration. Lecturers in LS seemed to use a larger number of dependent clause structures in order to direct the students or provide more information, most containing a subject and an action verb. The

most common sub-category in LS lectures is If-clause fragments which have been used 7% such as in: *if you want to, if you look at* and *if you have a*.

**Table 4.15 Proportional distribution of dependent clause fragments across the corpora**

| Sub-categories | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| 2a. 1st/2nd person pronoun + dependent clause fragment | 9(2.9%) | 15(4.5%) |
| 2b. WH-clause fragments | 8(2.5%) | 12(3.6%) |
| 2c. If-clause fragments | 11(3.5%) | 23(7.0%) |
| 2d. (verb+adjective) to-clause fragment | 19(6.1%) | 21(6.4%) |
| 2e. That-clause fragment | 7(2.2%) | 13(3.9%) |
| **Sub-total** | **54(14.3%)** | **84(25.4%)** |

(19) in terms of filtration clearance and I'm afraid the calculations that we can do from these parameters so ***if we look at*** an average man seventy kilograms and you can see a lot of us is water that's a huge amount *(LS-Medicine)*

(20) didactic story for our times and projective identifying oneself as biographer or a reader with another's story now ***if you think of*** those two terms i think neither of them allow for impartiality and objectivity *(AH-English)*

All the other sub-categories of dependent clause fragments were also higher in LS. The (verb+adjective) to-clause fragments were used with a similar rate, in LS lectures which comprise 6.4% of all bundles as opposed to 6.1% in AH lectures. Examples include: *to be able to, to talk about the* and *to look at the.*

(21) for instance Turner's paintings people couldn't read Turner's paintings originally but they came ***to be able to*** read them they came to be able to understand that they were of a storm at sea and so on *(AH-Philosophy)*

#### 4.3.4.3 The Comparison of Noun-Prepositional Phrase Fragments in the Corpora

The results also show that the third main category noun phrase and prepositional fragments were higher in AH lectures than in LS lectures. They were more dominant in AH lectures which comprised 47.6% of all bundles which means almost half of all bundles types are phrasal which consist of noun and prepositional phrase components such as in: *the end of the, a little bit of* and *the way in which*. However, lectures in LS accounted for 27.6% of all lexical bundles types Table 4.16. The high proportion of noun and prepositional phrase fragments in AH lectures might also have resulted from the expressive nature of the disciplines in this division that required the lecturers to use a greater variety of short phrases, including different nouns and prepositions, to best convey their disciplinary messages, as shown in the following examples from the corpus:

(22) the seminar in the next lecture I will look a little bit more on the issues of nationalism in Greece and ***the way in which*** archaeology was implicated in the growth of the Greek state as well as the actual impact *(AH-Classics)*

(23) we'll do that in week nine so what I've done instead today is i've picked out some more slides so at ***the end of the*** two lectures today we're going to have a couple of dozen slides to back up some of the things *(LS-Zoology)*

Table 4.16 indicates that noun phrase with of-phrase fragment by far registered the highest percentage of occurrence in the two disciplinary groups, with the AH instructors using this structure slightly more 18.7% than those of LS 11.6%. This structure was mainly used to quantify *(a little bit of),* qualify *(the nature of the)* or identify the topic being discussed *(at the start of)*, as in:

(24) is on allegory not just allegory in The Faerie Queen but allegory and symbolism in general and ***one of the things*** I'm trying to get you to see is that allegory in The Faerie Queene isn't just a sort of *(LS-English)*

(25) in terms of these letter forms is looking at approximately when and why they developed I'll talk a little bit about ***the nature of the*** evidence that we have concerning these display type specimens on the board *(AH-Typography)*

(26) you may ask where the leukocytes come from to do this test in the case of a live donor it's perfectly obvious you take ***a little bit of*** blood in the case of the cadaveric donor you use usually the spleen as a sort *(LS-Biological Sciences)*

Regarding the other sub-structures, lectures in AH also showed a higher dominance in using prepositional phrase fragments (19%) and other noun phrases expressions (5.5%):

(27) what this narrator is illustrating asking you to notice is that she is saying this is what childhood is while ***at the same time*** saying none of us can remember it and yet that's what she's claiming we do all *(LS-English)*

(28) collection of essays for example now let me look at an analysis of painting by Bryson and I will quote Bryson here but I have to give you ***a little bit more*** background before I can just launch into this quote *(LS-History)*

Prepositional phrase expressions seemed to be commonly used in LS 8.8% and noun phrase with post-modifier fragments 3.6% to make a logical relationship between the elements:

(29) we have to if you want to use less pesticides and use biological control we have to be able to accept that some ***of the things that*** we eat may not be absolutely perfect okay let's look at some slides before *(LS-Zoology)*

(30) because protons are the major metabolic end points as indeed is carbon dioxide so we can relate the control of blood to ***the way in which*** we are respiring if we're breathing deeply and rapidly that *(LS-Medicine)*

The two groups of lecturers made use of one lexical bundle *(as well as the)* that functioned as a comparative expression in presenting the subject materials. It would appear that this expressions are more frequently used in academic written discourse.

(31) we've been looking at the new processes of photography and lithography ***as well as the*** rebirth of the woodcut when it came back in the course of the nineteenth century as wood engraving next week *(AH-Typography)*

(32) to be as widely recognized as possible not just by within the business by the employees but also perhaps by your customers ***as well as the*** world at large and I mentioned N-V-Qs but there are other *(LS-Plant Sciences)*

**Table 4.16: Proportional distribution of NP +PP fragments in the corpora**

| **Sub-categories** | **Arts and Humanities** | **Life and Medical Sciences** |
|---|---|---|
| 3a. (connectors+) Noun phrase with of-phrase fragment | 58(18.7%) | 38(11.6%) |
| 3b. Noun phrase with other post-modifier fragment | 11(3.5%) | 12(3.6%) |
| 3c. Other noun phrase expressions | 17(5.5%) | 11(3.3%) |
| 3d. Prepositional phrase expressions | 59(19.0%) | 29(8.8%) |
| 3e. Comparative expressions | 3(0.9%) | 1(0.3%) |
| **Sub-total** | **148(47.6%)** | **91(27.6%)** |

In fact, it seems that the findings are in contrast with the earlier studies on academic English. Prior findings showed that academic speech primarily contained more lexical bundles with clause fragments and verb phrases, while academic writing reported to use

more bundles incorporating noun and prepositional phrase fragments (Biber & Conrad, 1999; Biber et al., 2004; Hyland, 2008).

All the sub-categories under noun phrase and prepositional fragments have been most popular among AH lectures. The difference in the use shows that the language of AH disciplines is more varied than those of LS, with more noun and preposition expressions. This corresponds to the idea that AH disciplines, like English and history, describe human-related issues such as diversity of behaviour. One way to describe or portray different characteristics of people is by using a range of noun and prepositional phrases. This fact allows for the assumption that the language of academic lectures in AH is more varied than the language of the academic lectures in LS.

## 4.4 Functional Analysis of Lexical Bundles in Academic Lectures

The following sections answer the second part of the research question 2:

2. How are the academic lectures in Arts and Humanities, and Life and Medical Sciences disciplines different or similar in terms of structural characteristics and discourse functions of the lexical bundle used?

In order to analyze the discourse functions of the identified lexical bundles, they were analyzed in their contexts to determine the discourse functions that they were trying to convey. Each of the 6609 tokens of the 636 types of lexical bundles were carefully checked and analyzed in their context to establish a clear picture of the communicative purpose that each of the lexical bundles conveys. If a bundle had more than one function, its primary function was categorized for the analysis by studying all the tokens in the context. For purpose of reliability, a second rater helped the researcher in the initial identification of communicative functions of the communicative function of the bundles

where necessary. After consultation, the final decisions were made regarding the function of target lexical bundles.

The comparison of lexical bundles and their discourse functions in the AH and LS lectures revealed a number of remarkable variations in the language use of the two corpora. The framework for analyzing discourse functions in this study was adopted from Biber et al.'s (2004) study. Biber et al.'s taxonomy has been designed on the basis of a broad corpus of written and spoken registers to include conversation, text books, academic writing and so on. The lexical bundles were distributed across the three main functional categories (stance bundles, discourse organizers and referential bundles) and their sub-categories in the two corpora. However, the focus of the present study is only on a spoken register in the form of academic lectures. Therefore, in order to fulfill the purpose of the present research and to best interpret discourse functions of some bundles in our corpus, some modifications had to be applied in their classification.

### 4.4.1 Revision of Taxonomy Proposed for the Functional Classifications of Lexical Bundles

In Biber et al.'s classification, the sub-category (text-deixis) under the referential expressions was changed to topic/subject/lecture reference as some bundles in this corpus referred to previously discussed topics in earlier parts of the lectures or prior sessions, rather than any text or course book (see Table 4.17). Finally, a fourth category was added (special conversational functions) and this category according to Biber et al. (2004) was found only in their conversation sub-corpus. It was devised with the aim of meeting the need of describing oral lectures as a specific genre investigated in the present study. The information in the following sections can be used to interpret how lexical bundles are used in the context of their discourse functions and together contribute to the shaping of the academic lectures in both corpora.

**Table 4.17: Revised version of Biber, et al., 2004 taxonomy of discourse functions of lexical bundles**

| **Functional categories** | **Sub-categories** | **Sample bundles** |
|---|---|---|
| **1. Stance expressions** | A. Epistemic stance | *I don't know if, the fact that the* |
| | B. Attitudinal/modality stance | |
| | B1)Desire | *if you want to,* |
| | B2)Obligation/ directive | *I want you to, it is important to,* |
| | B3) Intention/ Prediction | *I'm not going to, it's going to be,* |
| | B4) Ability | *to be able to, to can be used to,* |
| **2. Discourse organizers** | A. Topic introduction/focus | *what do you think* |
| | B. Topic elaboration/ clarification | *on the other hand,* |
| **3. Referential expressions** | A. Identification/ focus | *that's one of the,* |
| | B. Imprecision | *or something like that* |
| | C. Specification of attributes | |
| | C1) Quantity specification | *there's a lot of* |
| | C2) Tangible framing | *the size of the* |
| | C3) Intangible framing | *the nature of the* |
| | D. Time/ Place/ Text reference | |
| | D1) Place reference | *the united states and* |
| | D2) Time reference | *at the same time* |
| | **D3) Topic/subject/lecture reference** | ***as shown in figure*** |
| | D4) Multi-functional reference | *at the end of* |
| **4. Special Conversational Functions** | A. Politeness | *Thank you very much* |
| | B. Simple inquiry | *What are you doing* |
| | C. Reporting | *I said to him/her* |

### 4.4.2 Functional Classification of Lexical Bundles

All the 6609 tokens of the 636 types of lexical bundles which were identified in the corpora were analyzed in their contexts to arrive at the communicative functions that they conveyed in the lectures.

#### 4.4.2.1 Functional Classifications of Lexical Bundles in Arts and Humanities Lectures

As can be seen, Table 4.18 illustrates the distribution of lexical bundles across the three major functional categories (stance bundles, discourse organizers and referential bundles) and their sub-categories as well as the additional fourth functional category in the sub-corpus.

**Table 4.18: Distribution of lexical bundles across functional categories in Arts and Humanities**

| Functional categories | Sub-categories | No. % of bundles |
|---|---|---|
| **1. Stance expressions** | A. Epistemic stance | 17(5.5%) |
| | B. Attitudinal/modality stance | |
| | B1)Desire | 5(1.6%) |
| | B2)Obligation/ directive | 5(1.6%) |
| | B3) Intention/ Prediction | 25(8.0%) |
| | B4) Ability | 2(0.6%) |
| **Sub-total** | | **54(17.3%)** |
| **2. Discourse organizers** | A. Topic introduction/focus | 69(22.3%) |
| | B. Topic elaboration/ clarification | 37(11.9%) |
| **Sub-total** | | **106(34.2%)** |
| **3. Referential expressions** | A. Identification/ focus | 69(22.3%) |
| | B. Imprecision | 5(1.6%) |
| | C. Specification of attributes | |
| | C1) Quantity specification | 17(5.5%) |
| | C2) Tangible framing | -- |
| | C3) Intangible framing | 23(7.4%) |
| | D. Time/ Place/ Topic reference | |
| | D1) Place reference | 10(3.2%) |
| | D2) Time reference | 7(2.2%) |
| | D3) Topic/ subject/ lecture reference | 7(2.2%) |
| | D4) Multi-functional reference | 11(3.5%) |
| **Sub-total** | | **149(45.7%)** |
| **4. Special Conversational Functions** | A. Politeness | -- |
| | B. Simple inquiry | -- |
| | C. Reporting | -- |
| **Total** | | **309(99.1%)** |

As can be seen, referential bundles acted as a leading category which constituted 149 different bundle types, making 45.7% of the total bundles. Discourse organizer bundles were the second leading category with 106 strings and 34.2% of use and 54 (17.3%) bundles fell into the stance bundles, making it the third most common functional category. No special conversational functions were found. Appendix H shows the lexical bundle found in the sub-corpus of AH across the sub-categories of the main functions.

#### 4.4.2.2 Functional Classifications of Lexical Bundles in Life and Medical Sciences Lectures

As can be seen, Table 4.19 illustrates the distribution of lexical bundles across the three major functional categories and their sub-categories in the sub-corpus.

**Table 4.19: Distribution of lexical bundles across functional classifications in Life and Medical Sciences**

| Functional Categories | Sub-categories | No. % of bundles |
|---|---|---|
| **1. Stance expressions** | A. Epistemic stance | 16(4.8%) |
| | B. Attitudinal/modality stance | |
| | B1)Desire | 3(0.9%) |
| | B2)Obligation/ directive | 23(7.0%) |
| | B3) Intention/ Prediction | 48(14.6%) |
| | B4) Ability | 4(1.2%) |
| **Sub-total** | | **94(28.5%)** |
| **2. Discourse organizers** | A. Topic introduction/focus | 98(29.9%) |
| | B. Topic elaboration/ clarification | 20(6.1%) |
| **Sub-total** | | **118(36%)** |
| **3. Referential expressions** | A. Identification/ focus | 40(12.2%) |
| | B. Imprecision | 3(0.9%) |
| | C. Specification of attributes | |
| | C1) Quantity specification | 31(9.4%) |
| | C2) Tangible framing | 3(0.9%) |
| | C3) Intangible framing | 11(3.3%) |
| | D. Time/ Place/ Topic reference | |
| | D1) Place reference | 4(1.2%) |
| | D2) Time reference | 10(3.0%) |
| | D3) Topic/ subject/ lecture reference | 4(1.2%) |
| | D4) Multi-functional reference | 4(1.2%) |
| **Sub-total** | | **110(33.3%)** |
| **4. Special Conversational Functions** | A. Politeness | 1(0.3%) |
| | B. Simple inquiry | 3(0.9%) |
| | C. Reporting | 1(0.3%) |
| **Sub-total** | | **5(1.5%)** |
| **Total** | | **327(99.3%)** |

As can be seen, discourse organizer bundles acted as a leading category which constituted 118 different bundle types, making 36% of the total bundles. Referential bundles were the second leading category with 110 strings and 33.3% of use and 94 (28.5%) bundles fall into the stance bundles, making it the third most common functional category. The smallest number of bundle category was dedicated to special conversational

functions bundles with only 5 strings and 1.5% of use. Appendix I illustrates the lexical bundles in the LS sub-corpus which were used across the sub-categories of the main functions.

### 4.4.3 The Comparison between the Functions of Lexical Bundles in the Corpora

Analysis of the data revealed some outstanding similarities and differences between the corpora in relation to the communicative function of lexical bundles. Table 4.20 gives the numbers and percentages of the functional categories of lexical bundles in the corpora. In general, stance expressions accounted for a large portion of lexical bundles in academic lectures. This finding could support the earlier findings by previous studies which found stance bundles to be frequently used in a range of spoken registers such as conversation and classroom teaching (Biber et al., 2004; Biber & Barbieri, 2007). They found that in conversation, stance expressions accounted for over 60% of the bundles in their corpora. This similarity may highlight some shared features among spoken genres. In the same line, academic lectures commonly used self-reference markers using personal pronouns such *(I)* and *(We)* to express the lectures' stance towards a number of propositions. This basically shows that more transactional nature of academic lectures in this study, with the lecturer being the center of attention.

The comparison of lexical bundles and their functions in the lectures of AH and LS divulged a distinction in the language use of the two broad disciplinary groupings. From Table 4.20, there were some differences between the two disciplinary groupings in relation to the employments of lexical bundles in each functional category. One clear difference was the greater concentration of stance expressions in the LS lectures, a disposition which accounted for 28.5% of the bundles. It was followed by 17.3% in AH lectures. The greatest occurrence of stance bundles in LS signposts the LS lectures' highest inclination to show their presence in the lecture through using self-reference

markers. It might be due to the physical and laboratory-focused sense of the disciplines in this division which required more emphasis on the presence of the lecturer in the process of exchanging information than those in AH. This feature was, to some extent, shared among some disciplines of LS which dealt with experiments, measurements or procedures such as biological sciences. This fact is probably what made LS the leading sub-corpus in the use of stance bundles. On the other hand, the lower occurrence of stance bundles in AH can be sourced from the less empiricist nature of the disciplines in this academic division which minimize the presence of the lecturer in the delivery of information.

**Table 4.20: Frequency of lexical bundles across functional categories in the corpora**

| | **Functions** | **Arts and Humanities** | **Life and Medical Sciences** |
|---|---|---|---|
| **1** | Stance expressions | 54 (17.3%) | 94 (28.5%) |
| **2** | Discourse organizers | 106 (34.2%) | 118 (36%) |
| **3** | Referential bundles | 149 (45.7%) | 110 (33.3%) |
| **4** | Special conversational functions | -- | 5 (1.5%) |
| | **Total** | **309 (99.1%)** | **327 (99.3%)** |

A similar pattern of use was found regarding the use of discourse organizer bundles in LS that reflected the higher tendency of the LS lectures to organize their language by providing a link between prior and forthcoming discourse through formulaic expressions. The plausible interpretation for such phenomena can be the explanatory-based nature of the hard expressions in LS which could have driven the lectures to rely more on the use of lexical bundles in the delivery of subject materials. LS is mostly based on procedures and experimental steps, dealing with methods and equipment used, to establish their discourse more coherently by providing a link between prior and forthcoming information. This consequently might facilitate better interpretation of the procedures in the experiment. These procedural steps were also seen in a few disciplines of LS which

dealt with agricultural botany and health care. That may be the reason why LS lectures also exercised the choice of a great use of discourse organizers. A similar pattern was found in AH with the percentage of 34.2%. These lectures were found to be more explanatory and less empirical than soft sciences in which experimental steps and procedures are not the norm in their disciplines.

Another major difference between both disciplinary divisions was that AH were dominated by the use of referential bundles, which comprised almost half of all the bundles. AH makes direct references to different types of entities in their lectures through the use of the formulaic chunks. Fields in AH require more interpretation and contain evaluative patterns of argument (Hyland, 2008) in the process of importing their content. The lecturers tend to introduce a variety of concept, and specify some important attributes about them or make reference to different parts of the lectures in order to draw attention of students and to link together points in the argument to build a cohesive message. Whereas, LS with the more objective nature, use less referential bundles to forward the points made in the lectures.

Results in Table 4.20 report that there are no expressions that fall under the category of special conversational functions in the AH. On the other hand, in the LS fields, lecturers engage their audience by providing the students with necessary bundles presented in their previous lectures at the onset of a lecture. Another function is the use of lexical bundles that serve to end a lecture by introducing the topic that will be delivered in the next lecture.

The next sub-sections will compare the sub-categories of the main functions of the lexical bundles across the lectures of the two corpora with examples extracted from each corpus. The information in these sections and tables can reveal how lexical bundles and

their discourse functions can be different across disciplinary groupings, thereby showing the different choices made to organize the speech patterns accordingly.

#### 4.4.3.1 Comparison of Stance Bundles in the Corpora

Analysis of the stance bundles in the corpus of the two corpora showed that, in general, these expressions were used in the academic lectures to frame the speakers' attitudes and assessment towards a number of current events. They were of two types, personal or impersonal. While personal bundles obviously dealt with lecturers' stance (*I don't know how* and *I think if you)*, impersonal bundles conveyed the same meaning but there is a shift from the direct to the more indirect form, such as *(are likely to be)*. Findings in Table 4.21 showed that most of the stance bundles in this study were personal, highlighting the involvement of the lecturers and taking a personal stance in the delivery of the academic lectures. The bundles included impersonal reference whereby the lecturers overtly addressed themselves or the audiences. Such use of reference seems to be characteristic of academic spoken discourse in which the presence of the speaker and the referring to "self" suggest a greater emphasis of the presence of the lecturer when compared to that of written discourse. It had been observed that, authors of written discourse preferred not to situate a personal stance in their claims as they connect with their readers by presenting the content in a more objective manner (Hyland, 2002). Stance bundles had two main functions or meanings in academic lectures: epistemic and attitudinal/modality stance bundles.

**Table 4.21: Proportional distribution of stance bundles in the corpora**

| Sub-categories | Arts and Humanities | Life and medical sciences |
|---|---|---|
| A. Epistemic stance bundles | 17(5.5%) | 16(4.8%) |
| B. Attitudinal | 37(11.8%) | 78(23.7) |
| B1. Desire | 5(1.6%) | 3(0.9%) |
| B2. Obligation/ directive | 5(1.6%) | 23(7.0%) |
| B3. Intention/ Prediction | 25(8.0%) | 48(14.6%) |
| B4. Ability | 2(0.6%) | 4(1.2%) |
| **Total** | **54(17.3%)** | **94(28.5%)** |

The two groups of lecturers made use of stance bundles in a fair amount to show a variety of propositions such as certainty, uncertainty, ability and intention and direction (17.3% in AH compared to 28.5% in LS). There was only slight differences between both corpora with regards to epistemic stance expressions. The result seems to be in line with the previous study by Biber et al. (2004), who found that most epistemic stance bundles in spoken registers, such as conversation and classroom teaching, are used to show personal rather than impersonal engagement.

**(A)** ***Epistemic stance bundles***

The epistemic stance bundles in this study showed the lecturers' degree of knowledge towards the coming information. Most of the epistemic bundles were personal expressions (*i don't know if, i don't think it's* and *i thought it was)*, using the first person pronoun *(I),* signaling the lecturers' preference to involve a personal evaluation or opinion about the state of knowledge given before or after the bundle, as in shown in the following examples:

(33) sludge language is a way of world-watching there was another fascinating example that turned up in the Guardian some while back again ***I don't know if*** this is true but it said it was in the Guardian Japanese *(AH-English)*

(34) I don't think the people of Framingham realize that and what we know about micro albuminuria ***I don't know how*** they had the foresight in the early sixties to start looking for very small amounts of protein in urine *(LS-Medicine)*

(35) the sort of energy and exchange and net radiation things like the microclimate modification radiation is that they've all got a part in understanding and predicting what's going on ***and I think that's*** quite important *(LS-Plant sciences)*

(36) the Russian Revolution or whatever but think of it how is he setting up that he or she setting up the argument another angle I would say and ***I think this is*** a crucial one look at the language and above all look at the (*AH-History*)

(37) I don't yet know whether anything changed because ***I don't know what*** to expect from a disease what I would expect disease progression in a field where nothing changed to be like okay we are very used *(LS-Agricultural Botany)*

This type of usage appears to be in contrast with the previous findings found by Conrad (1996) and Reppen (2004), whereby they suggested that in written academic discourse (e.g., textbooks), it was preferred not to use personal stance in their arguments or claims in order to keep text objectivity (Hyland, 2002).

Most of the epistemic bundles in each corpus were personal, disciplinary lecturers' preference to involve themselves and their students in evaluating the propositions that follow the lexical bundle. There were a number of variations between the two groups in relation to the use of these bundles. Table 4.22 shows the distribution of the epistemic bundles in both disciplinary groupings. It is usual for AH disciplines to focus on people by default people's behaviour is varied. Thus, this leads to a greater use of lexical bundles to support claims. On the other hand, LS disciplines deal with presenting of the lecture materials, thus the lecturers make use of many bundles that show uncertainty.

**Table 4.22: Proportional distribution of epistemic stance bundles in the corpora**

| Function | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| A. Epistemic stance bundles | 17 (5.5%) | 16 (4.8%) |

Epistemic bundles express a state of uncertainty. These bundles in the two corpora reflect the lexical bundles which are mainly used to keep the lecturer's personal opinion.

Another variation in the use of such bundles could be identified to be particular to the type of disciplinary divisions. For example, LS relied on some bundles that include the modal verb that expresses possibility such as *would,* or the modal *may* which bundles were used to show only possibility, they occurred in both the LS and AH lectures as follows:

(38) and that ***there would be a*** perfectly legitimate a perfectly good kind of undergraduate essay that might have no very personal original slant to it but which is a very lucid a very just you know helpful summary of the debate *(AH-English)*

(39) other sexually transmitted diseases seems to be bad news there is a suggestion although it's never been qualified that ***there may be a*** genetic component to susceptibility to progression into AIDS now that's *(LS-Biological Sciences)*

(40) those results I put up before are really sort of in experiments I think we're probably using measure photosynthesis and ***it may be that*** plants adapt to the rates to the levels of C-O_ two and reduce their rates *(LS-Plant Sciences)*

(41) the key to understanding women's history is in accepting painful though ***it may be that*** it is the history of the majority of the human race so Virginia Woolf I think toys with the reader trying to tease out this recognition *(AH-English)*

In contrast, there were fewer impersonal epistemic stance in each disciplinary group. The impersonal bundles *(are more likely to, the fact that the* and *it seems to me)* occurred in both AH and LS, which reflect possibility or to show that something mentioned is not a fact but a personal opinion. Some impersonal bundles reflect a degree of certainty to keep the lecturers personal judgment out of the lecturers and to help maintain an objective tone (Conrad, 1996; Rappen, 2004) as seen in the following examples:

(42) innocence but like Henry James' good American Twain too had eaten of the tree of knowledge and ***it seems to me*** that as Huckleberry Finn took shape Twain found it less possible to blinker out the facts *(AH-English)*

(43) American jazz music but is very often composed by French composers and what we can measure the success they had by ***the fact that the*** theme tune in both Grisbi and Rififi become big hits in the mid-fifties *(AH-Film and Television studies)*

(44) to have low income we also know that black and minority ethnic households ***are more likely to*** have income less than fifty per cent of average income than white households and these are figures again *(LS-Medicine)*

As can be seen from the above examples, stance bundles were found particularly at the beginning of university lectures as suggested by Cortes and Csomay (2007). Also these bundles were found to be used by lecturers to negotiate class management issues and to guide class participation in the middle of the lecture (Cortes & Csomay, 2007).

**(B)** ***Attitudinal/ modality stance bundles***

The two groups of lecturers used a variety of attitudinal/modality expressions in order to express their feeling and attitude towards the occasions happening according to its

related propositions: desire, obligation, directive, intention/prediction, and ability (Table 4.23).

**Table 4.23: Proportional distribution of attitudinal/ modality bundles in the corpora**

| Function | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| B1)Desire | 5(1.6%) | 3(0.9%) |
| B2)Obligation/ directive | 5(1.6%) | 23(7.0%) |
| B3) Intention/ Prediction | 25(8.0%) | 48(14.6%) |
| B4) Ability | 2(0.6%) | 4(1.2%) |
| **Total** | **54(17.3%)** | **94(28.5%)** |

**(B1)** ***Desire bundles***

Desire bundles express only personal expressions, which demonstrate personal wishes or inquires about another participant's desires *(*e.g., *if you want to, I just want to* and *I don't want to)*. This function was showed to be used similarly by AH and LS lecturers (Table 4.24).

**Table 4.24: Desire bundles in the corpora**

| Function | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| B1. Desire | 5(1.6%) | 3(0.9%) |

Lecturers in LS used desire bundles to clarify the instruction or task (example 45), while these bundles were used mostly as a point of departure in the AH lectures, as in example 46:

(45) to increase cardiac output that is you increase the stroke volume and the heart rate and you increase total peripheral resistance you constrict the arterioles so ***<u>if you want to</u>*** raise blood pressure you have *(LS-Animal and Microbial Sciences)*

(46) to investigate how did we come to think in that way now that's a route Thompson didn't want to go down and ***I don't want to*** go any further down it now we will explore that much more next term *(AH-History)*

(47) a big percentage increase that's this part of the loop the baroreflex increasing the heart rate to try and keep your blood pressure up ***i just want to*** point out here that it's not just your arterioles *(LS-Animal and Microbial Sciences)*

(48) draws you into thinking in terms of the natural the eternal the uncivilized and it's important to think about how those are suggested finally ***i just want to*** say something very briefly about Rome *(AH-film and television studies)*

A few lexical bundles that express personal desire in academic lectures are also used sometimes to initiate new topics. These bundles have two functions either a desire bundle or as a discourse organizer bundles such as the following bundles *(what I want to)*.

(49) I'll come back to the question in a moment so ***what I want to*** do is just describe to you some of the things that most people would agree are true about neo-realistic films and as I say once again *(AH-Film and Television Studies)*

**(B2)** ***Obligation/ directive bundles***

Obligations or directives are understood as personal expressions of stance, directing the listener to carry out actions that the speaker wants to have completed, or expressing predictions of future events that do not involve any participation of the speaker *(*e.g., *you don't have to, you need to know* and *you look at the)*.

Results in Table 4.25 showed that obligation/directive bundles were markedly more popular in LS, with the lecturers dedicating more than four times as many lexical bundles

to this function than those in AH. These bundles accounted for 7% occurrence in LS and 1.6% in AH. This higher occurrence was more evident in the use of obligation markers which seems to be typical of LS fields, in which the lecturer was required to direct the listeners to carry out an experiment or to complete a required task. Most of these bundles are personal stance bundles and they differ from other personal bundles in that they have a second person pronoun *(you).* For example the bundle *(you need to know)* was used to stress the necessity of the proposition (example 51).

**Table 4.25: Obligation/ directive bundles in the corpora**

| Function | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| B2)Obligation/ directive | 5(1.6%) | 23(7.0%) |

(50) if you didn't do the research that knowledge would not be available you could not get it so ***you have to do*** the research to generate new knowledge audit is more a retrospective looking at current practice *(LS-Biological Sciences)*

(51) and depending on the age of the textbook or the age of the clinician they might alternate between A-V-P or A-D-H so ***you need to know*** both of them and the water transport in the collecting duct *(LS-Medicine)*

(52) all these different people have said and that's why it's a perfectly legitimate kind of essay to do but ***you can see that*** there is the danger of reading things or taking notes and they all go into a kind of *(AH-English)*

Most of the bundles found under this subcategory functioned as obligation rather than direction in the in LS. This might have resulted from the physicality nature of the disciplines in this science that required the lecturers to directly or indirectly force the students to do some actions as an obligation.

In some examples, these bundles included verb of desire with a first person pronoun *(I),* directly conveying the speaker's desire that the addressee carried out some action, and thus functioning as directives:

(53) at the end of this talk I'm going to give you a list of fifty facts and ***I want you to*** analyse these facts with some degree of caution okay tell me a little bit now about some more specific causes of kidney stones *(LS-Medicine)*

In contrast, the directing function of these bundles was more dominate in AH lectures. The AH sub-corpus used some directive lexical bundles *(*e.g., *i want you to, you might want to, you have to remember* and *you look at the).* These bundles were personal expressions, using second person pronoun *(you).*

(54) proper handling is sometimes known by the term scientific history you'll find that in some of your books when you see that though ***you have to remember*** that's actually a translation of the German word *(AH-History)*

Some examples of the directive bundles were used in a very indirect way as an example the bundle *(you might want to)* that was found in AH sub-corpus:

(55) to the elements and become strong men and she completely takes the position apart so ***you might want to*** look at that if you're interested in that kind of issue I think it's a text which does that *(AH-English)*

Some directive bundles were used also as topic introduction in both AH and LS. Such bundles can be called multi-functional bundles *(*e.g., *have a look at):*

(56) all the traditional view of what was happening within Britain and then we can ***have a look at*** the problems created by the new evidence so we go back to commius who establish himself in the territory *(AH-Classics)*

(57) we're just not too much detail in the lecture because you get them in case studies to actually look at them in more depth so if we ***have a look at*** urine production first off then so this is you've got *(LS-Medicine)*

Only one impersonal obligation/directive bundle was found in both groups of lectures. This bundle was used with no personal pronoun at all *(it has to be),* even though they still clearly directed the listeners to carry out some action:

(58) loss of function as I've already said this is irreversible cannot be reversed and once the organ has failed ***it has to be*** removed and you start again though I've said that it is probably not an immunological response *(LS-Biological Sciences)*

(59) the child is posited as not innocent as fallen already so the child has to aspire to spiritual redemption ***it has to be*** saved okay so this is not a vision of a child who is innocent this is an idea of childhood *(AH-English)*

Three of these bundles: *so you have to, you have to do* and *we have to do* were specific to LS and were not found in the corpus of AH. In the following examples from two lecture in biological sciences, the lecturer was explaining a procedure and steps.

(60) if you measure them at different times of the day you get different numbers ***so you have to*** measure them at specific time in the day because there's a natural biorhythm of C-D- four level cells in the bloodstream *(LS-Biological Sciences)*

(61) if you didn't do the research that knowledge would not be available you could not get it so ***you have to do*** the research to generate new knowledge audit is more a retrospective looking at current practice *(LS-Biological Sciences)*

Directive bundles were also used in LS lectures to indicate the lecturer's stance of guiding students to a particular procedure/step or to point to a figure:

(62) the proportion of individuals who have incomes which are below fifty per cent of average income and ***you can see here*** from these figures in nineteen-seventy-nine we had about nine per cent *(LS-Health and Social Sciences)*

The bundle *(if you look at)* was used in LS because it includes empirical studies which deal with the sense of touch, hearing, movement, change and people's life. It was used as a directive bundle in LS to guide students towards the applied tools (example 63), and as a topic introducing bundle in AH (example 64).

(63) you producing okay so that probably would be the best thing if you had a look at that and I'll try and remember to bring some but ***if you look at*** those posters they're exactly the sort of thing you know *(LS-Agricultural Botany)*

(64) the painter called er Lusieri who was er Elgin's er agent that's L-U-S-I-E-R- he was in charge of Elgin's operations on the ground in Athens er ***if you look at*** his letters to his patron er it's quite interesting *(AH-Classics)*

**(B3)** ***Intention/prediction bundles***

With regard to intention/prediction bundles, findings indicated that these expressions were also more preferred by LS lecturers (Table 4.26).

**Table 4.26: intention/ prediction bundles in the corpora**

| Function | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| B3) Intention/ Prediction | 25(8.0%) | 48(14.6%) |

In general, these bundles were used more personally (*what we're going to* and *and we're going to*) then impersonal (*is going to be*) in both corpora. In LS lectures these bundles were used as a way to explain certain procedures/ steps taken, as in the following examples, in which the lecturers use the personal pronoun (*we)* to help engage the

students in the task and make them feel that they are part of the lesson. The bundles in examples 65 and 66 were used as introducing a topic.

(65) it also is important for bones we've talked about this before ***and we're going to*** so whether the calcium stays in your bones or is released in the circulation vitamin D has a role in and it also regulates *(LS-Medicine)*

(66) to start looking for very small amounts of protein in the urine but they did and this is a very clever idea the reason I emphasize the first two stages is that I think that if ***we're going to do*** anything about diabetic *(LS-Medicine)*

On the other hand, the impersonal bundles mainly expressed future prediction about incoming incidents and did not necessarily include the speakers' decision, as in:

(67) if it is true that removes their freedom but the reward of virtue certainly is ***not going to be*** happiness in all circumstances we hope it will in many but we need to go right back to that distinction *(AH-Philosophy)*

Data analysis revealed some variations between the two groups of lectures regarding the deployment of these bundles. A closer look at Table 4.26 indicates that these bundles occurred in LS with 14.6% more than in AH 8%. Experimental disciplines are more stage-based, with the experiments and results normally explained in steps. Therefore, it appears that LS lectures used more intention/ prediction bundles to explicitly inform audiences about their intentions or possible prediction about the next tasks and steps in the experimental procedure. These bundles were used in the LS lectures to elucidate a process that involve steps:

(68) another question that you might try and ask as a plant breeder is what ***is going to be*** the spectrum of pets and diseases in seven years' time and can assure you these do move and do change quite a lot over time *(LS-Agricultural Botany)*

(69) the stomach is actually a very acid environment so viruses which ***are going to be*** denatured at low PH are going to be destroyed in the stomach so enteroviruses are particularly tough and resist the virus *(LS-Biological sciences)*

On the other hand, in AH sub-corpus, personal intention bundles also expressed future actions such as in the bundle (*and i'm going to*), which is used to point to something important. One important difference in the use of prediction/intention bundles was that the first personal pronoun *(I)* was most common in AH, while *(we)* was mostly used in LS lectures.

(70) to be lecturing on Orlando A Biography today ***and i'm going to*** concentrate on three main aspects today the biographical aspect the whole issue of parody and issues about relationship to history *(AH-English)*

(71) we dismiss the claims of empiricist historians to produce objective facts about the past ***I'm not going to*** even begin to try to offer you any kind of definitive answer to that question because what from one direction *(AH-History)*

Shared bundles such as the bundle *(you're going to be)* was used to refer to speaker's prediction of proposition in LS lectures (example 72). In AH the same bundle was used as an obligation force as in (example 73). This suggests that disciplinary lectures have their own way of organizing ideas.

(72) put on something with no explanation people are not going to get anything from it I mean you've got okay ***you're going to be*** there to answer questions but the poster has got to be free-standing *(LS-Agricultural Botany)*

(73) to speak both in terms of explaining the children's motives not going on there they were just told if you kill your brother ***you're going to be*** killed too don't behave like that here presently they began to *(AH-English)*

Another difference between corpora was with the type of lexical bundles used in each corpus to serve this function. In this regard, LS relied more on the expressions of joint actions that included the use of the collective first person pronoun *(we)* while planning the intentions. Examples include bundles such as: *and we're going to, we're not going to, what we're going to, if we're going to, that we're going to, we're going to have* and *we're going to start* which only occurred in the sub-corpus of LS. This way, the lecturers attempted to invite the students as participants in the tasks and activities. This would have made the learners become more aware of the empirical tasks and procedures introduced in the discipline.

However, there were some variations in the way the disciplinary lectures used bundles of this type. In view of these variations, several intention/ prediction bundles had a dual function in one group but not in the other. As such, most of these bundles were also used as a trigger to initiate a discussion or topic in some disciplines under investigation. For example:

(74) a nerve impulse it's got to go one way yes now ***what we're going to*** find is that this will reduce let's have a look at it on the few title spots of blue will appear I need something to measure out one *(LS-Mathematics)*

(75) disadvantaged communities but let's first of all look at inequalities in access to primary health care and ***what we're going to*** see is a bit of a mixed bag the reviewers er of r- of that have looked at er the breadth of *(LS-Medicine)*

(76) this type of activity to a large extent okay I'm going to drop the lights a little bit here cause I'm ***going to show you*** a few slides of what actually want to looked at er on a on an audit around this nursery *(LS-Plant Sciences)*

**(B4)** ***Ability bundles***

All ability bundles that occurred in AH and LS lectures were personal (Table 4.27).

**Table 4.27: Ability bundles in the corpora**

| Function | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| B4) Ability | 2(0.6%) | 4(1.2%) |

The lexical bundles *to be able to* were the only one found in both AH and LS lectures. The following examples show that some examples occur together with directive bundles to identify skills and tasks that students should complete:

(77) what tells what story but we want you ***to be able to*** read in the detail for yourselves if the Faerie Queen were no more than what the useful lecture summaries make it would neither have been worth writing *(AH- English)*

(78) therefore we make progress with plant breeding so we need ***to be able to*** produce and release genetic variation and I'm sure you're all know very well what is the most obvious and straightforward way to produce *(LS-Agricultural Botany)*

#### 4.4.3.2 Comparison of Discourse Organizing Bundles in the Corpora

Discourse organizer bundles played an important role in building cohesion and were used to organize the lecturers' messages as they "reflect relationships between prior and coming discourse" (Biber et al., 2004, p. 384). In general, lecturers in LS tended to use discourse organizer bundles slightly more than those of AH (36% compared to 34.2% respectively). Table 4.28 shows the distribution of discourse organizer and its sub-categories across the two disciplinary divisions.

There are two major sub-categories included under discourse organizing bundles: topic introduction/focus and topic elaboration/clarification. Topic introduction bundles provide obvious signals that a new topic is being introduced. Many of these are expressions of intention or desire, but quite a number of them have a more specialized function, that of announcing the intention to begin a new topic. The second major sub category of text organizing bundles is topic elaboration or clarification bundles.

**Table 4.28: Proportional distribution of discourse organizing bundles in the corpora**

| Sub-categories | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| A. Topic introduction/focus | 69(22.3%) | 98(29.9%) |
| B. Topic elaboration/clarification | 37(11.9%) | 20(6.1%) |
| **Total** | **106(34.2%)** | **118(36%)** |

**(A)** ***Topic introduction/ focus bundles***

Regarding the first sub-category, Table 4.29 shows that topic introduction/ focus bundles account for 29.9% in LS which is higher than those that occur in AH lectures with 22.3%. It might partly be due to the presence of more difficult and novel content in the lectures of LS which resulted in using more pauses. Consequently, the lecturers had to start a new sentence in order to avoid ambiguity in the message and promote

understanding among the students. The other possible reason was the application of some instruments or tools which required the lecturers to use some topic initiating bundles in order to attract the students' attention and invite their participation towards the instrument.

**Table 4.29: Topic introduction/focus bundles in the corpora**

| Sub-categories | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| A. Topic introduction/focus | 69(22.3%) | 98(29.9%) |

Topic introduction bundles such as: *if you look at, what I want to* and *going to talk about* were mostly found at the beginning of the sentence to initiate a lecture or a new topic or signal topic change as in:

(79) now ***what i want to*** do today is to look at another case study cause we were looking last time at the Imperial Way Via dell at the centre of Roman connecting the colosseum up to the Victor Emmanuel monument *(AH-Classics)*

(80) what i'm ***going to talk about*** briefly now is rhinoviruses a small R-N-A virus which as you know I hope causes colds common colds inhalation of aerosols bearing this virus results in infection and this *(LS-Biological Sciences)*

(81) looking at their growers suppliers for example I'm talking about growers themselves perhaps ***going to look at*** their suppliers because if schemes like customers that anything's in order because they will *(LS-Plant Sciences)*

In addition to these variations, a closer look at the analysis of bundles showed that each group of lectures resorted to specific types of four-word lexical bundles to serve topic introduction/ focus function in their presentation. Such a phenomenon signals the fact that apart from the shared bundles to introduce a new topic; some lexical bundles

were disciplinary-specific. For example, one way to introduce a topic in LS was by using the bundles: *now i'm going to, and we're going to, what we're going to, i'm going to show* and *we're going to start.*

(82) come back and apparently I'm still giving the lecture I don't know how that happened but here I am and ***what I'm going to*** talk about is Marxism er and Matrix historiography since Marx linking up *(AH-History)*

The topic introduction bundles *a bit more about* and *talk a little bit* occurred in AH and LS, containing the quantity markers *(bit, little)* to help minimize the scope of the lecture. These examples basically show how the language choice of the lectures varies across the two groups, that is, in order to function as a trigger to start a new discussion, each group of lectures make use of specific set of chunks.

(83) what we've observed under our null hypothesis has about a five per cent chance of occurring I'll ***talk a little bit*** more about how about how we choose a cut-off point p-values a bit later on third example *(LS-Statistics)*

(84) you know the programme's more important than the outcome so these are ways in which you might er er think ***a little bit about*** the language changing in your own time see we're not really ultimately *(AH-English)*

The discourse organizers category accounted for the most lexical bundles with a dual function in the whole corpus, especially in the case of topic introduction/focus sub-category in AH and LS corpora. For instance, the directive bundle *if you look at* served a topic introduction function in some lectures in AH and LS:

(85) but that this period of tyranny had not been sufficient to squash the nationalist impulses of the Americans so ***if you look at*** The Jamaica Letter I think you can you'll see good examples of this sort *(AH-Comparative American Studies)*

The bundles *(I want to talk + want to talk about)* comes together to form one longer bundle *(I want to talk about)* in LS lectures as in:

(86) to get going the last three issues of nature have had big steps on the way what ***I want to talk*** about however is something different I want to say that many different creatures prokaryotes *(LS-Mathematics)*

(87) to get going the last three issues of nature have had big steps on the way what I ***want to talk about*** however is something different I want to say that many different creatures prokaryotes *(LS-Mathematics)*

Lectures in LS used the bundle *(to look at the)* with a directive function (example 88), while in AH this bundle was used to introduce topics as in (example 89):

(88) toxic so what they did with the synthetic pyrethroids was ***to look at the*** chemical structure of pyrethrum which is a chemic- a naturally occurring chemical structure of pyrethrum *(LS-Zoology)*

(89) what I want to do is actually ***to look at the*** literary record today and to see whether in fact er there are dangers within it that we need to to bear in mind when we look at Agricola as a historical *(AH-Classics)*

The bundle *(if we look at)* was used as a transition maker directing the students towards the material objects embedded in the lecture (example 90), while in AH it was used to shift a topic.

(90) things like er increases in benefits and pensions just to show you why er access to income in kind is important ***if we look at*** this slide you can see that er this is an analysis by Shaw et al er and *(LS-Medicine)*

According to Biber et al., (2004), the use of some topic introduction bundles could result in syntactic blends (p.104). In example (91), the bundle *(if we look at)* occurred in the middle of the sentence following the connectors *(so, well, but, and)* and functioned as a trigger for the proposition following the bundle.

Another interesting finding was the use of other bundles which included the verbs *see* and *look* as in the bundles *if you look at, you can see that* and *so you can see.* Studying the context in which these bundles revealed that the AH sub-corpus, many of these bundles primarily functioned as topic introduction and secondly as directive. However, this pattern of use was opposite to the use in LS that their main function was to draw the students' attention towards the proposition following the bundle in LS. Such fluctuation in discourse function of some lexical bundles reflects the fact that once we move from less empirical to more empirical fields, some expressions are assigned more discipline-specific function.

(91) well what evidence is there actually for Pericles what what evidence is there for him being for him being considered important in his own day well ***if we looks at*** Thucydides actually associates Pericles *(AH-Classics)*

(92) they're deficient elderly people in calcium from what we would normally count as normal levels so ***if we look at*** plasma concentration or plasma calcium that's what I've told you that's *(LS- Medicine)*

In contrast, the main function of the bundle *have a look at* was used as a directive in AH and LS (see examples 93, 94), whereas it was primarily used by LS lectures to

indicate a change in the topic, mostly preceded by the word *let's* as in the following examples found in the corpus:

(93) and mouthing off every so often about this that er and the others so let's ***have a look at*** some new kinds of history read Marc Bloch's Historian's Craft that's what you got to do on the that's the next semester *(AH-History)*

(94) now what we're going to find is that this will reduce let's ***have a look at*** it on the a few little spots of blue will appear oh I need something to measure out one ml excuse me now when I put one ml *(LS-Mathematics)*

**(B)** ***Topic elaboration/ clarification bundles***

The second sub-category is topic elaboration/ clarification, where AH lectures accounted for more types of those bundles than in LS lectures, as can be seen in Table 4.30. AH were reported to use nearly twice as many topic elaboration/clarification bundles as LS (11.9% compared to 6.1% respectively). This can reflect the greater inclination of AH lecturers to the use of formulaic expressions to make a connection between prior and forthcoming discourse. Such interest may result from the need for more discussion of ideas in this discipline, which calls for logical connections to ease students' comprehension. That could account for why topic elaboration/clarification bundles were more favoured by the AH lecturers.

**Table 4.30: Topic elaboration/ clarification bundles in the corpora**

| Sub-categories | Arts and Humanities | Life and medical sciences |
|---|---|---|
| B. Topic elaboration/clarification | 37(11.9%) | 20(6.1%) |

For example, the bundles (*to do with the, on the other hand)* were used to represent the distinction between coming and prior information in the following examples:

(95) stable unified self is dissolved in the text ***on the other hand*** there's clearly a fascination with an aristocratic family whose lineage stretched back to Elizabethan times and with the whole notion of the *(AH-English)*

(96) I want to agree with you what the topic is so that A it's not impossible for you to do for being too wide but ***on the other hand*** there's enough chance of you finding some literature and something to say about it *(LS-Agricultural Botany)*

(97) developing world and that's probably got simply ***to do with the*** fact that people are generally more malnourished in the developing world so in other words you know their general state of health is *(LS-Biological Sciences)*

Table 4.30 presents information on the distribution of topic elaboration/ clarification bundles in the lectures of both corpora. As shown in the table, there was no difference between both groups of lectures regarding the use of lexical bundles, except that there was a difference in the range of use. However, there was an apparent variation between both corpora of lectures regarding the type of bundles used to carry out this function. LS lecturers used the bundles *(what i mean by, nothing to do with* and *I want to say),* but this was not used in AH. Another example was the use of the bundles *(we can say that, that is to say* and *on the one hand)* which only occurred in AH. Such examples can portray the diversity in the selection of lexical bundles among lectures belonging to different academic fields.

Some lexical bundles were used to clarify a term or a topic specific to a discipline, *(what I mean by)* in the following example taken from the LS sub-corpus:

(98) some sets of alleles tend to stick together okay and occur at a higher frequency than others ***what I mean by*** the term of haplotype I think you'll probably intuitively understand is you have a series of loci along *(LS-Biological Sciences)*

#### 4.4.3.3 Comparison of Referential Bundles in the Corpora

Referential bundles are used to create logical connections between different propositions discussed in the lectures (Biber et al., 1999). Nesi and Basturkmen (2006) finds that the appositive associations in the lectures are signalled by referential bundles *(*e.g., *one of the things, of the things that* and *a lot of the).* That is, referential bundles generally identify an entity or single out some particular attributes of an entity as especially important. There are four subcategories included under referential bundles performing four main functions: identification/focus, imprecision, specification of attributes and time/place/text reference.

The two disciplinary divisions also record some marked similarities and differences. One major difference is the higher concentration of referential bundles in the AH lectures, amounting to 45.7% of all the bundles types. However, in LS lectures, this category comprises one-third of all the bundles types (33.3%). The detailed information about the sub-functions with the percentages of use is given in Table 4.31.

**Table 4.31: Proportional distribution of referential bundles in the corpora**

| **Sub-categories** | **Arts and Humanities** | **Life and Medical Sciences** |
|---|---|---|
| A. Identification/ focus | 69(22.3%) | 40(12.2%) |
| B. Imprecision | 5(1.6%) | 3(0.9%) |
| C. Specification of attributes | | |
| C1. Quantity specification | 17(5.5%) | 31(9.4%) |
| C2. Tangible framing | - | 3(0.9%) |
| C3. Intangible framing | 23(7.4%) | 11(3.3%) |
| D. Time/ Place/ Topic reference | | |
| D1. Place reference | 10(3.2%) | 4(1.2%) |
| D2. Time reference | 7(2.2%) | 10(3.0%) |
| D3. Topic/subject/ lecture reference | 7(2.2%) | 4(1.2%) |
| D4. Multi-functional reference | 11(3.5%) | 4(1.2%) |
| **Total** | **149(45.7%)** | **110(33.3%)** |

**(A)** ***Identification/focus bundles***

As regards to the first sub-category, identification/focus bundles, focus on the noun phrase following the bundle as especially important. In many cases, identification/focus bundles also have a discourse organizing function. Identification/focus bundles in AH lectures show a higher percentage of use (22.3%), leading to the conclusion that these lectures comprise more expressive topics that need to be identified in the form of giving definition or characterizing some important features, such as in the shared bundles *and one of the, one of the reasons, is one of the, one of the things, and this is the,* and *is a kind of* which have the highest rate of occurrences. This type of bundles is used mostly in AH lectures (see Table 4.32).

**Table 4.32: Identification/ focus bundles in the corpora**

| Sub-categories | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| A. Identification/ focus | 69(22.3%) | 40(12.2%) |

Some of the examples that use these bundles are shown below:

(99) in reason is the slave of the passions very famous quote from Hume reason ***is a sort of*** secondary thing the passions the emotions the feelings are what really govern us and not reason is something that comes *(AH-philosophy)*

(100) Montaillou Village Occitan on the other hand is about a small peasant village and its inhabitants at a turbulent moment in the history of the region and ***one of the things*** it talks about is the is the Countess who *(AH-English)*

(101) these were set out as booty capitalism that is the robber barons pariah capitalism now this ***is a kind of*** commercial activity which he identified with forms of money lending and again sets this discuses *(AH-History)*

In contrast, these types of bundles are found to be used less often in LS fields. These findings are in contrast with the findings of previous studies by Kashiha (2014). Previous research shows that the hard sciences use more bundles of this sub-category than soft sciences. These bundles deal with a variety of abstract and concrete entities. They may also refer to materials or instruments which need to be identified and explained through a number of specification expressions *(*e.g., *and this is the, is one of the, one of the things* and *one of the most)*. The following lines show the use of these bundles in context:

(102) the thing I need to tell you about is if we go back one thing I need to stress and ***it's one of the*** mysteries of nucleic acid structure that always mystifies people if we look at the structure as we've draw it here *(LS-Animal and Microbial Sciences)*

(103) only once photosynthesis isn't something that happened only once therefore ***and this is the*** argument if you ran evolution again on this planet you'd get photosynthesis you'd get life because it's downhill to chemistry *(LS-Mathematics)*

In some cases, this bundle *(one of the most)* also serves as a topic introduction function. In AH, several notions and concepts are required to be identified or defined by the lecturers. In LS, the lexical bundle is used *(one of the most)* primarily to identify the noun which comes before or after the bundle as in examples 104 to 107:

(104) the sort of problems that you can have environmental problems that you can have now the most or ***one of the most*** effective herbicides that's a pesticide for killing other plants is a chemical called two-four-five *(LS-Zoology)*

(105) it's been a topic of interest really to lots of educational writers probably ***one of the most*** famous ones is John Dewey as you can see by the black and white photo he was born in eighteen fifty nine *(LS-Medicine)*

Biber et al., (2004) found that in classroom teaching, the bundle *(those of you who)* is used to connect to a sub-group of students, similarly in lectures use was found in LS and AH.

(106) you'll see this with Jane Eyre I'm going to look as in a moment for ***those of you who*** are reading Jane Eyre or have read Jane Eyre er Villette Lucky Snow any I narrated novel has a problem *(AH-English)*

(107) I guess as being something like blanket stitch okay in out round the back okay ***those of you who*** think I've got the nomenclature of the stitch wrong can come up and tell me later but I assure you there *(LS-Animal and Microbial Sciences)*

**(B)** ***Imprecision bundles***

Imprecision bundles indicate imprecise reference *(*e.g., *or something like that* and *and things like that)* which is the least frequently occurring bundles in both corpora (see Table 4.33).

**Table 4.33: Imprecision bundles in the corpora**

| Sub-categories | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| B. Imprecision | 5(1.6%) | 3(0.9%) |

This suggests that the lecturers are almost sure of the exactness of the references they are using, or there are no supplementary references of the same type. There were no important differences between both sub-corpora regarding these bundles. These bundles are used almost similarly, as can be seen in the following contexts (108-110) in both corpora.

(108) we'd have a vase of flowers and maybe in front of the vase of flowers you might have a selection of fruit maybe some dead game ***or something like that*** as well the kind of still life idea when we look at one *(AH-Philosophy)*

(109) they adverse effect quality pollutants don't get so diluted ***and things like that*** in effect quality and fairly obviously more intense winter rainfall can give increased flooding et cetera and there's evidence *(LS-plant Sciences)*

(110) a quick reminder of what a hypothesis test is we set up our hypothesis which is to quantify our belief about say an incident rate ratio ***or something like that*** and we calculate the probability of what we've observed *(LS-Statistics)*

**(C)** ***Specification of attributes***

Regarding the third sub-category of referential bundles that specify amounts or quantities, the LS registers a slightly higher proportion of the specification of attribute bundles than AH (12.9% compared to 13.6%). There are three functions under the specification of attributes: quantity specification, tangible framing bundles and intangible framing bundles as presented in Table 4.34.

**Table 4.34: Proportional distribution of specification of attributes bundles in the corpora**

| Sub-categories | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| C. Specification of attributes | | |
| C1. Quantity specification | 17(5.5%) | 31(9.4%) |
| C2. Tangible framing | - | 3(0.9%) |
| C3. Intangible framing | 23(7.4%) | 11(3.3%) |
| **Total** | **40 (12.9%)** | **45 (13.6%)** |

In support for the high occurrence of specification attribute bundles, Hyland (2008, p. 16) finds 50% of text-oriented bundles in social sciences texts "to frame agreement

by highlighting connections, specifying cases and pointing to limitations". However, here there are few variations between the corpora as explained in the following sections.

**(C1)** ***Quantity specification bundles***

Quantity specification bundles function as measuring the amount, number or quantity of the attributes in the lectures and used with high frequency in the lectures of the two disciplines (e.g., *a lot of the, a little bit more, the rest of the, a little bit of* and *quite a lot of).* Quantity specification bundles are the most frequent sub-function in the LS sub-corpus (9.4% compared with 5.5% in AH) as shown in Table 4.35. The higher occurrence in LS could be explained based on the necessity for more description and elucidation of the concepts in the two corpora. LS is somehow descriptive in terms of quantity and thus the use of such bundles is important.

**Table 4.35: Quantity specification bundles in the corpora**

| Sub-categories | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| C1. Quantity specification | 17(5.5%) | 31(9.4%) |

Lines 111 -116 below shows some of the examples of how quantity specification bundles are used in both the LS and AH corpora:

(111) an average height or whatever but you should remember that it's not a value not a range in which ninety-five ***per cent of the*** observations lie and you can illustrate that quite easily if we split up the data *(LS-Statistics)*

(112) the top right hand corner of one of the small slit-like windows there and I'm looking at that now think of how on sees ***the rest of the*** building as you look at that do you understand what I'm getting at here *(AH-Philosophy)*

(113) the current football stadium reduces the number of people in there the other key thing to think about although ***a lot of the*** artistic style has a reference back to antiquity the way it's drawn and the way *(AH-Classics)*

However, there are some variations in relation to the manifestation of its sub-functions. For example, the quantity bundle (*a little bit about)* also functions more specifically as a topic introducer in LS:

(114) we're trying to achieve today in terms of these letter forms is looking at approximately when and why they developed I'll talk ***a little bit about*** the nature of the evidence that we have concerning these display *(AH-Typography)*

The bundle *(a little bit more)* is used in AH specifically to emphasize the amount of knowledge delivery to pre-empt unfavourable response from the audience but it is used in LS mainly to specify quantity of amount:

(115) all this now er that's to put a very broad frame of reference let's sort of move in then ***a little bit more*** specifically by varying stages into our topic one thing that I think would be very useful for you *(AH-English)*

(116) before I finish the sort of key quality management systems I just want to say ***a little bit more*** about benchmarking because it is a slightly different issue to anything we've talked about before *(LS-Plant Sciences)*

**(C2)** ***Tangible framing***

Tangible framing bundles describe the form and the size of the following noun (e.g., *the size of the* and *this part of the*). There is one major difference found between AH and LS lectures, that is, there is no lexical bundle found in AH lectures that functions as

tangible framing bundles. There are only 3 occurrences of this function in the LS lectures *(*e.g., *this part of the, the size of the* and *the structure of the)* as shown in Table 4.36.

**Table 4.36: Tangible framing bundles in the corpora**

| Sub-categories | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| C2. Tangible framing bundles | - | 3(0.9%) |

The following lines (117-118) show how tangible framing bundles are used in the LS sub-corpus:

(117) the protein bends and the proline is at a key area here that forms an interaction between ***this part of the*** protein and this part so the moral to remember all the time is that just occasional amino acids at key points *(LS-Agricultural Botany)*

(118) if you remember I mentioned that although the kidney has an extremely high blood flow for ***the size of the*** organ you only get five to ten per cent of this actually going into the medulla so that's *(LS-Medicine)*

**(C3)** ***Intangible framing bundles***

Unlike tangible framing bundles, intangible framing bundles identify abstract characteristics. A closer look at Table 4.37 reveals that the two disciplinary divisions record a low rate of use regarding intangible bundles. However, intangible framing bundles are more favoured by AH lecturers and comprises 7.4% of the bundle types, compared with only 3.3% in LS.

**Table 4.37: Intangible framing bundles in the corpora**

| Sub-categories | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| C3. Intangible framing bundles | 23(7.4%) | 11(3.3%) |

These bundles tend to function as building some cohesion for discussion in the lectures. Unlike LS which deals with physical and real objects, the main focus of AH is on conceptual properties of entities. For example, the bundles *(in the context of, in terms of the,* and *the nature of the)* which are only used in AH lectures, contribute to the explanation of contexts under which the topics are being discussed.

The following examples (119-122) present how these bundles are used in context in both the LS and AH corpora:

(119) the reason why that's problem is because one of the consequences of pregnancy is naturally immunosuppress so ***as a result of*** becoming pregnant a woman will naturally immunosuppress that is *(LS-Biological Sciences)*

(120) the continually transforming effects of industrialization and the what this has brought about in not just ***in terms of the*** economic development of individual countries but in terms of the economic development *(AH-History)*

In some cases, abstract specifying bundles were used to establish logical relationships in a text. Such bundles pinpoint some abstract features about the entities discussed in the disciplinary lecture *(*e.g., *the nature of the* and *the context of the)* as in:

(121) and isn't just a matter of correctness it's a matter of understanding something about ***the nature of the*** language itself so if you have a good etymological dictionary and think of it in terms of *(AH-English)*

(122) last point to note is that statistically significant result is not necessarily a clinically important one again this depends on ***the context of the*** problem that we're dealing with one that I've consulted on recently *(LS-Statistics)*

**D. *Time/ Place/ Topic reference***

The fourth sub-category, time/place/topic reference, refers to particular places, times or location in the text. These bundles reveal some similarities and minor disciplinary variations between the lectures. AH uses more of these sub-category of referential references than LS corpus (11.1% compared to 6.6% in LS). As shown in Table 4.38, AH lecturers use the higher proportion of these bundles. Time reference bundles in LS are the only type that are higher in percentage than those in AH.

Biber et al., (2004) find several referential bundles in their corpus of textbooks, academic prose, classroom teaching and conversation, to refer to particular places, time and location. The only difference between the present corpus and that of Biber et al. (2004) is that no bundle is found in academic lectures to refer to particular location in text. This corpus is oral presenting of lectures therefore, the lectures were found not to refer to the text while lecturing.

**Table 4.38: Proportional distribution of time/place/topic reference bundles in the corpora**

| Sub-categories | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| D. Time/ Place/ Topic reference | | |
| D1. Place reference | 10(3.2%) | 4(1.2%) |
| D2. Time reference | 7(2.2%) | 10(3.0%) |
| D3. Topic/subject/ lecture reference | 7(2.2%) | 4(1.2%) |
| D4. Multi-functional reference | 11(3.5%) | 4(1.2%) |
| **Total** | **35(11.1%)** | **22(6.6%)** |

**(D1) *Place reference***

Lecturers in AH use place reference bundles more than lecturers in the LS sub-corpus with 3.2% and 1.2% respectively, as shown in Table 4.39.

**Table 4.39: Place reference bundles in the corpora**

| Sub-categories | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| D1. Place reference | 10(3.2%) | 4(1.2%) |

(123) the Bible the Bishop's Bible authorized by Elizabeth's government which has a woodcut frontispiece and the frontispiece represents Elizabeth ***at the top of*** the page sitting in her throne actually with her *(AH-English)*

(124) here we have a map showing broadly you can look at the one ***in fort of you*** if you prefer the contours of the colonial viceroyalties in seventeen-eighty and you can see up here here's the viceroyalty *(AH-Comparative American Studies)*

(125) that really isn't true even with dialysis the mortality's fifty per cent so the word failure let's have the lady in pink ***at the back there*** tell me about the word failure and what do you think the patients *(LS-Medicine)*

(126) you have to look for key conserved amino acids and we knew from this analysis that clearly the conserved er histidines in ***the centre of the*** protein were some of the most functionally interesting *(LS-Agricultural Botany)*

**(D2)** ***Time reference***

LS lecturers used time reference bundles slightly more bundles than those used by AH lecturers, as in Table 4.40, especially in some topics of medicine and biological sciences lectures which had to do with periods of time:

**Table 4.40: Time reference bundles in the corpora**

| Sub-categories | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| D2. Time reference | 7(2.2%) | 10(3.0%) |

(127) actually something quite important you can reduce ***over a period of*** months the amount of drugs that are used and this is telling us that in fact that the patient is becoming acclimatized if you like become adapted *(LS-Biological Sciences)*

(128) give them an angiogram which is nephrotoxic or tomorrow we may want to give them T-P-N or N-G-P we may dialyse people for rather odd reasons but ***most of the time*** they are only two absolute reasons for dialysis *(LS-Medicine)*

(129) it suggests perhaps fatherland it's also feminine it's a feminine noun so it somehow suggests motherland ***at the same time*** and indeed people sometimes like to talk even more all inclusively not just about *(LS-Comparative American Studies)*

**(D3)** ***Topic/subject/ lecture reference***

Topic/subject/ lecture reference the revised sub-category, demonstrated a similar scale of use in the two disciplinary divisions (2.2% in AH and 1.2% in LS) (see Table 4.41). These bundles were used to make somehow direct or indirect reference to previously-mentioned issues in the lectures.

**Table 4.41: Topic/subject/lecture reference bundles in the corpora**

| Sub-categories | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| D3. Topic/subject/ lecture reference | 7(2.2%) | 4(1.2%) |

These expressions were normally used in the introduction section to initiate the lecture by providing overviews to previously related topics, as in examples from LS lectures*(*e.g., *go back to the, come back to that, i'll come back to* and *we'll come back to)*.

(130) it's really would be impossible to test every single pesticide on every single potential species in the wild and ways of getting around this problem ***I'll come back to*** later on when we talk about ecotoxicology *(LS-Zoology)*

(131) I've told you the prediction of how it might have changed during evolutionary time but what did it really look like and i ***go back to the*** comment that said this protein was a multimeric protein *(LS-Agricultural Botany)*

These expressions were used to propose a discussion at a later session, they were usually found towards the end of the disciplinary lectures, together with the introducing of the future topics. One way to show this function was the use of the expressions such as *(come back)* in the bundle *(we'll come back to)* which occurred in the corpora. The bundles *(we'll come back to* and *come back to that)* can form only one bundle *(we'll come back to that)*. This bundle was only found in LS sub-corpus.

(132) malaria the spraying of pesticides managed to contain malaria outbreak quite successfully for a number of years unfortunately ***we'll come back to*** this later but unfortunately of course mosquitoes *(LS-Zoology)*

Example (132) showed that this function is important in facilitating the opening and closing of a cohesive lecture.

In contrast, lecturers utilized preview bundles to make reference to the topics which would be explained later in the following session or the next lectures. These bundles were mostly found in the body section of the lectures, using the first person singular/plural *(I/ we)* to engage the students. Examples from the AH lectures *(i'll come back to, to come back to, come back to that, come back to this, we'll be looking at, i was talking about* and *i said at the):*

(133) the two images very much point to the two bodies if you like Elizabeth and ***i'll come back to*** this idea in a moment Elizabeth's public monarchical self the sword in its scabbard it's there to be drawn *(AH-English)*

(134) pretty simple pretty much all secondary qualities are going to count as simple perceptions the immediate feel the immediate but again this is controversial so we're going ***to come back to*** this later *(AH-Philosophy)*

(135) he puts it the past is not a child's box of letters with which we can spell any word we please and in fact as ***I said at the*** beginning the rhetorical strategy of the book is to set that sceptical trap only *(AH-History)*

(136) I say spin doctors of course they weren't professional P-R people but they knew how to manipulate the public pretty effectively and ***we'll be looking at*** some of the ways in which they did this on the *(AH-English)*

Similarly, bundles such as: *I was talking about* can be used in the beginning of a lecture as a reminder or in the middle of the discussion just to recap what has been said. It was interesting to note that in giving an overview of the previous discussion, the use of *(I)* was common in this type.

(137) they were doing so that was just was incredible nobody had actually seen a computer at all anyway when ***I was talking about*** this sort of thing then I used a different example which was of you know *(AH-Philosophy)*

**D4.** ***Multi-functional reference***

Multi-functional reference were used more often in AH lecturers than in LS (3.5% compared with 1.2% in AH), as in Table 4.42. These bundles were used to refer to a specific period of time (example 138), particular place (example 139) or topic depending on the context (example 140), such as the bundle *(at the end of)* which had three different functions:

**Table 4.42: Multi-functional reference bundles in the corpora**

| Sub-categories | Arts and Humanities | Life and medical sciences |
|---|---|---|
| D4. Multi-functional reference | 11(3.5%) | 4(1.2%) |

(138) the first generation here we just have one density ***at the end of*** the second generation however in both cases we will have two densities okay in conservative three of them will be light one of them *(LS-Animal and Microbial Sciences)*

(139) some guys put ***at the end of*** their patient information sheet thank you for taking part in this research and that's an obsolute no it's considered coercive so it's things like that you need to know now *(LS-Biological sciences)*

(140) some do survive they retain their specificity the important point about that is it means that ***at the end of*** the infection there are many more cells specific for that pathogen than there are the individual *(LS-Biological Sciences)*

As for the disciplinary difference, the bundle *at the end of* served two functions in AH lectures, a place function in (example 141), while it made reference to time (example 142):

(141) this kind of satire this parody this critique and one of the things she parodies is precisely our tendency to frame historical periods so ***at the end of*** chapter four she writes and I've given you that you that as the eight *(AH-English)*

(142) organizing it so further details can be got from her and I'll leave these with the handouts ***at the end of*** the lecture okay so any questions on that just yet it's in June the twenty-second to the twenty-second to the *(AH-Philosophy)*

Another example is the bundle *(go back to the)*, it was used only in the LS lectures and it had two purposes: either to refer to place (example 143) or to direct audiences by refereeing to the slides or examples previously discussed (example 144):

(143) a farmer went out and collected seed grew them and they did relatively well the chances are he'd ***go back to the*** same place to collect seed or use seed he'd grow okay if a neighbour *(LS- Agricultural Botany)*

(144) the prediction of how it might have changed during evolutionary time but what did it really look like and I ***go back to the*** comment that said this protein was a multimeric protein it had different subunits *(LS-Agricultural Botany)*

#### 4.4.3.4 Comparison of Special Conversational Functions in the Corpora

As shown in Table 4.43, no lexical bundles are found to function as special conversational functions in AH lectures. The three sub-functions are manifested in different sections of the lectures. In total, lecturers in LS appear to have a high interest in using these bundles.

**Table 4.43: Proportional distribution of special conversational functions in the corpora**

| Sub-categories | Arts and Humanities | Life and Medical Sciences |
|---|---|---|
| A. Politeness | - | 1(0.3%) |
| B. Simple inquiry | - | 3(0.9%) |
| C. Reporting | - | 1(0.3%) |
| **Total** | **-** | **5(1.5%)** |

As shown in Table 4.43, there is only one example of the politeness expression (e.g., *thank you very much)* in LS lectures. This expression is mostly used in the body section

and closing sections of the lectures to announce the end of the lectures. In some cases, the lecturers used this bundle to thank the audience for their presentations in class or for attending the class (example 145). In other cases, this bundle is found at the beginning or end of the lectures (example 146).

(145) will you be writing that in your portfolio I will certainly ***thank you very much*** for coming and thank you for everybody that has contributed and prepared for today and as I say particularly for *(LS-Medicine)*

(146) I think quite rightly you have a break for the next 3 or 4 minutes just to stretch your legs outside and then please come back in the lecture theatre in the next five minutes ***thank you very much*** *(LS-Medicine)*

There are only 3 examples of simple inquiry expressions: *what do you think, does that make sense* and *are there any questions* and these expressions are used in the body of the lectures to engage students in the lectures and to keep up with the flow of information.

The yes-no question *(are there any questions)* is used in order to ascertain whether the audience gets the nuance of what has been discussed as in the following example:

(147) to normal and that's due to this tissue swelling ***are there any questions*** about this so far okay well these are two cases where things go wrong for the Starling balance but there we can cope with it the body can *(LS-Animal and Microbial Sciences)*

The common example of wh-question expressions is: *what do you think* which is placed usually near the end of the discussion after a long description to elicit information and make sure whether the audience has followed the discussion.

(148) it seems sustainable to you do you think that generally what we talked about could practically do in an enterprise of twenty or so people ***what do you think*** do you believe in it what we've been talking about *(LS-Plant Sciences)*

In contrast, the findings of Biber et al. (2004), show that the bundle *what do you think* serves as a topic introduction in classroom teaching and conversation, but in this study, this bundle is found to be used mainly to stimulate the audience's reaction towards the discussed topic by evoking their opinion. This shows that each bundle can vary in function depending on the context in which it occurs.

This expression *I said to you* is normally used in the introduction section to initiate the lecture by providing overviews to previously related topics by direct reporting through the use of this bundle. The bundle *i said to you* is used in the lecture opening with topic-reference that functions in LS sub-corpus:

(149) I finished yesterday by talking about stage one which is the acute or primary infection stage and ***i said to you*** that during this period that people were seroconverting they were they'd been infected they were *(LS-Biological Sciences)*

## 4.5 Summary of the Chapter

This study aims to identify and analyze the structural and functional characteristics of lexical bundles in university lectures from two broad disciplinary divisions. On the whole, it is found that lexical bundles are frequently used in the academic lectures of both disciplinary divisions. The findings of this study are in line with previous studies like Biber et al., (2004), Biber and Barbieri (2007), Kashiha and Chan (2013, 2014a, 2014b), and Kashiha (2015) in analyzing structures and functions of lexical bundles. It could be inferred that certain bundles are distinct to the nature of the disciplines. Regarding the structural classifications, the analysis reveals that most lexical bundles used in AH

lectures are noun-phrase + prepositional fragments, including prepositional phrase fragments, and noun phrase with of-phrase fragments. On the other hand, lectures in LS use more verb-phrase fragments such as 1st + 2nd personal pronoun + verb phrase fragments, and verb phrase with non-passive verbs. Concerning the discourse functions that lexical bundles serve, the findings show that stance bundles are more common in LS lectures, while discourse organizing bundles are used in a similar way in both disciplinary divisions. Almost half of the lexical bundles in AH lectures serves referential purposes as well as they are very common in LS lectures. With regards to the last category (special conversational functions), it is found only in LS lectures. However, even though the number of bundles that carry special conversational functions used in LS is small, they serve important communicative functions.

Also, the findings of this study suggest that there is a close relationship between forms and functions of lexical bundles. For instance, lexical bundles that incorporate noun phrase + prepositional phrase fragments are usually bound with referential bundles. Moreover, most of the verb-phrase fragments are used to express stance and discourse organizing functions. These findings are in line with previous studies Strunkyt and Jurkūnait (2008), Kashiha and Chan (2013, 2014a, 2014b), Kashiha (2015), that there is a link between structural categories and discourse functions. General conclusions from the findings of the present study will be further presented and discussed in Chapter 5, along with the implications and recommendations for further research.

# CHAPTER 5: CONCLUSION

## 5.1 Summary of Key Findings

The focus of this study on lexical bundles as a target linguistic feature has been motivated by the significant role it plays in fluent linguistic production and successful language learning (Hyland, 2008). Lexical bundles are one of the main sources of coherence and "important building blocks in discourse" (Biber and Barbieri, 2007, p. 270). The type of formulaic chuncks frequently used in the language of native speakers is a good way to empower the speaking fluency by second language learners and become more native-like in the target language, as these expressions have are important role in organizing the discourse. This study employs a corpus-based approach to investigate the extent to which lexical bundles as cohesion building blocks in discourse could be used differently in academic speech across disciplinary divisions.

In meeting the objectives, this study has identified and compared the frequency, structure and discourse function of the most frequently occurring lexical bundles in academic lectures of two broad disciplinary groupings namely Arts and Humanities, and Life and Medical Sciences by applying a corpus-based approach. The need to conduct such research of academic lectures was initially motivated by the need to understand the use of lexical bundles in academic lectures, especially those bundles that differ between broad disciplinary divisions in order to facilitate their use by teachers and learners. In line with previous studies (Biber et al., 2004; Biber & Barbieri, 2007; Kashiha & Chan, 2013, 2014a, 2014b; Kashiha, 2015), this study was able to describe the bundle use in terms of the more prevalent communicative functions that are associated with them. In this way, some properties can be arrived at to inform the pedagogical approaches that can be adopted based on the findings.

From the analysis of lexical bundles in this study, a summary of the results is presented in this section in order to have an overview of the study. In relation to the frequency of lexical bundles used in the two broad disciplinary divisions, the following conclusions were drawn. First, lectures in LS used the slightly larger number of lexical bundles types and tokens compared to AH. There were 327 different four-word lexical bundles in LS, while AH used 309 types of bundles. A similar pattern of use was found regarding the tokens or the total occurrence of the bundles in two corpora. LS lectures made use of 3,388 individual cases of lexical bundles compared with 3,221 tokens used by AH. Lecturers in LS rely slightly more on the use of lexical bundles to organize their language and avoid ambiguity. More lexical bundles could be needed in LS disciplines because these disciplines rely on formulaic patterns in order to link their reported results and to forward the message to audiences. These findings were in line with the findings of previous studies by Hyland (2008) who also found that hard sciences such as Engineering and some pure and applied sciences as Biology showed a greater tendency to use lexical bundles because the subject matter is more technical. Similar findings were also found by Kashiha and Chan (2013, 2014a, 2014b) and Kashiha (2015) where the frequency of occurrence of 4-word lexical bundles in hard sciences were larger than those found in soft sciences.

Among all the target bundles retrieved for analysis in the study, 105 bundles were found to be shared between the two groups of lectures, singling the commonalities of oral academic language showing the reliance on specific sets of lexical bundles across the disciplines. From the analysis, such variations in the number of shared bundles between disciplines simply point to visible differences in the language choices of the disciplinary lectures. The lexical bundles identified were compared with the list identified by Biber et al. (2004) in his sub-corpus of classroom teaching. It was found that there were a total of 78 shared bundles between LS sub-corpus and Biber's sub-corpus of classroom teaching,

whereas 49 shared bundles were found between AH sub-corpus and Biber's classroom teaching sub-corpus. This finding indicates that lexical bundles were frequently used in university lectures and accounted for a larger proportion of academic lecturers' discourse.

It was also disclosed that a number of lexical bundles were specific to only one discipline and were not found in the other disciplines. In view of the findings, LS contained a slightly larger number of these bundles, with 218 types compared with 204 in AH. The varied nature of disciplinary divisions has resulted in a large number of specific bundles in each corpus. Along with the large stock of shared chunks, each group of lectures appeared to rely on specific bundles in each corpus in order to get their message across. For instance, the top five most frequently occurring lexical bundles in AH sub-corpus were *on the one hand, the idea of the, all the rest of, as a kind of,* and *and all the rest,* and in the LS sub-corpus these are *i'm going to talk, and things like that, thank you very much, per cent of people,* and *if you have a.*

Regarding the variations in the structural use of lexical bundles across the corpora, the following conclusions were attained. Both groups of lectures showed different tendencies towards the construction of lexical bundles. In view of this, the largest number of lexical bundles in AH was those constructed with noun and prepositional phrase, while lexical bundles in the form of verb phrase fragments were most common in LS. Kashiha (2015) also reported the same results, that is, social sciences used more lexical bundles that were formed with noun and prepositional phrase, while Life and Medical Sciences used the most number of verb phrases to form lexical bundles. This diversity in structural selection could also have resulted from the differing divisions' backgrounds. AH were seen to be more descriptive in nature, with many concepts and notions required to be presented in the lectures. This could have led to the lectures using more combinations of nouns and prepositions to show the relationship between the terms used and in introducing such

concepts, they could be weaved more cohesively for the audience' understanding. However, lectures in LS, could be more governed by the need to use other main structures such as verb-phrases fragments. Such lexical bundles are more suited for the expression of content that are generally considered to be more empirical in nature, and much of it deals with tools, objects, measurements and human subjects. The emphasis on experiments could have driven the lectures to report the results through the use of more dependent clauses *(*e.g., *you can see how* and *we can see here)*, and these were used to connect to the use of visual aids during the lectures.

The obtained results imposed the following conclusions regarding the similarities and differences in discourse functions of lexical bundles in the two broad disciplinary groupings. In general, stance bundles were the most common function of lexical bundles in a range of spoken registers such as conversation and classroom teaching (Biber et al, 2004 and Biber & Barbieri, 2007) as in the frequent use of self-reference markers to reflect speakers' stance towards proposition mode. From the analysis, functional distribution of bundles in both sub-corpora revealed a variation in language use of the two broad disciplinary groupings. Lecturers in the two sub-corpora resorted to different lexical bundles in order to perform communicative functions related to their fields. In general, stance bundles were used in academic lectures to reflect the lecturer's attitude and evaluation towards a number of proposition that relate to attitudes such as desire, intention, prediction, direction/obligation and ability. The findings of this study showed that LS dedicated a higher proportion of stance bundles, while less attention was given to them by AH. This divulges the importance of lecturers' attendance in LS sub-corpus to dealing with things more physical and experiment-based and thus may require a greater explicit presence of the lecturer. On the other hand, the use of stance bundles in AH was seen as being less common in the corpus and their subject matter was less empirical and therefore, may result in downplaying the presence of the lecturer. Same results were found

in the studies by Kashiha and Chan (2013) and Kashiha (2015) who found that stance bundles were also more common in hard sciences than in soft sciences. In contrast, Kashiha and Chan (2014b) found that soft sciences used a greater proportion of stance bundles than in hard sciences.

Discourse organizing bundles were used in academic lectures to frame a relationship between prior and forthcoming information and served two main functions, either as: topic introduction/ focus or topic elaboration/clarification bundles. Regarding the use of discourse organizing bundles, findings indicated that this function of lexical bundles was used in a similar rate in both LS and AH lectures. Disciplines in pure and applied sciences deal with steps and procedures in conducting experiments and this could have driven the lecturers in LS to use slightly more discourse organizing bundles to establish the necessary links in previously mentioned content to the one that is to be introduced or to clarify a topic in the lecture, so that the audience can better interpret the results and procedures of experiments. Some disciplines in LS, such as agricultural botany or those related to health care issues were also found to have procedural step. The more descriptive and less experimental nature of soft sciences such as AH would not see the need to resort to lexical bundles that perform this function. The same pattern of use was found in the study by Kashiha (2015), and findings indicated that this function was more found in hard sciences, whereas the least inclination towards these bundles were detected to soft sciences.

In the corpus of academic lectures, referential bundles were the largest number of bundles found in both corpora. These bundles were used to help the lectures give reference, identify entities or specify more important functions about them in discussing topics related to disciplinary divisions. Lectures in AH used the larger number of referential bundles, with almost half of the total bundles found to be dedicated to enact

this function. In AH fields, lectures are required to give more reference to different parts of the oral content using lexical bundles to guide audience through lectures. In addition, referential lexical bundles occurred in LS sub-corpus at higher rate too which show how specifying attributes and giving reference are considered more important in some disciplines of LS such as zoology and botany which are more likely based on interpretative reasoning. Similarly, Kashiha (2015) found that social sciences used more than half of the total bundles dedicated to enact this function.

No marked differences were found between corpora concerning the use of the last functional category, special conversational functions, which was created to account specifically for the features that characterize the genre of oral academic lectures. Although occurring with far less frequency compared to the other function, they are nonetheless significant. Though these bundles were found in the LS lectures, no lexical bundles were found in AH lectures. These features were found to explain the structural cohesiveness of lectures which could adopt important communicative purposes such as giving an overview of the previous lectures at the beginning of a new lecture or introducing the topic of the next lecture at the end of the class to give closure to the communicative acts. Thus the newly added category, special conversational functions, in the form of lexical bundles can provide valuable contribution to the description of structure of a genre, that of academic oral speech. This gives a holistic framework for such speech analysis and provide future research with an analytical approach that might be insightful and useful.

There was a relationship between structural and functional categories. That is, sometimes there was a one to one relationship between the form and function of the bundles in that a particular structure was used to serve certain functions. For instance, noun and prepositional phrase fragments were the most common structures in the

production of referential expressions *(*e.g., *in the case of).* Verb phrases and dependent clause fragments, on the other hand, were mostly used to construct stance expressions *(*e.g., *I think that is)* and discourse organizers *(*e.g., *I would like to).* As for the special conversational functions (e.g., *thank you very much*) this was mostly used in the introduction section of the lectures. In addition to the above findings regarding the discourse function of lexical bundles, it was also disclosed that a number of bundles served more than one function in one sub-corpus but not in the other.

Overall, the data reported in this study was compared to other studies on spoken discourse (Kashiha & Chan, 2013, 2014a, 2014b and Kashiha, 2015). The results of the present study was compared to other studies such as Nesi and Basturkmen (2006) who explored lexical bundles of university lectures also from the BASE corpus. Their main focus was on exploring the cohesive role of spoken discourse in relation to the use of lexical bundles. However, their study did not focus primarily on highlighting disciplinary variations when explaining the discourse functions of the lexical bundles. Other studies investigated lexical bundles from another perspective whereby researchers like Biber and Barbieri (2007), and Cortes (2004) investigated lexical bundles by comparing their use in written and spoken registers. Other researchers such as Hyland (2008) explored lexical bundles in the writings of university students from different disciplines.

## 5.2 Limitations and Suggestions for Pedagogical Applications

Before discussing the pedagogical implications of the findings of this study, it is deemed necessary to present concisely the limitations of this research. The first limitation of this study is its corpus size. The study only analyzed 40 academic lectures from each disciplinary division which may not be adequate for generalizing the reported findings. Further studies with larger corpora could be conducted to have a more insightful picture

of the use of lexical bundles and their structures and discourse functions in academic disciplinary lectures.

The number of academic disciplines focused on is also another issue leading to the limitations of this study. Only two disciplinary divisions were covered. To come up with more generalizable findings in relation to cross-disciplinary variations, it seems logical to focus on more than two academic disciplinary groupings to yield more comprehensive and generalizable results.

The limitations also point to the need to further work on analytical framework used for lexical bundles. More modifications and justifications might be needed to improve on the framework. Thus, further research could be motivated towards this area to have a more detailed analytical framework to provide a solid foundation in the analysis of lexical bundles.

Further research could extend the area and look at the use of lexical bundles as an essential base for the problems EFL/ ESL learners have with the recognition of discourse types and functions of these frequently occurring expression in the language. An authentic corpus can be build based on lectures presented in L2 contexts such as those found in Malaysia compared with those of native speakers to articulate similarities and differences and see how L2 lecturers resort to native-like fluency using the formulaic language used by native speakers. Also other types of multi-word expressions such as collocations could be used as the subject for other related studies. An analysis of other linguistic features such as cohesive and metadiscourse markers can be compared with lexical bundles in academic lectures across disciplines. Research can also be more discrete in terms of the investigation in that stance alone could be focused on as the ubiquitous future in academic oral speech to see how it is conveyed not only through lexical bundles but other ways as well. This could also apply to discourse organizer bundles and referentials.

## 5.3 Implications

This study has compared the use of lexical bundles in the academic lectures of two broad disciplinary groupings in order to highlight and describe how lexical bundles were deployed with functional properties in different disciplinary lectures by native speaker lecturers The most important pedagogical implications that result from the findings of this research is the variation between frequency, structures and discourse functions of lexical bundles in the lectures of the two fields. Students who study in those divisions could be familiarized with the functions of the bundles in academic speech in university setting through conscious learning if they have problems in following lectures. In other words, these students through English for Academic purposes (EAP) courses may learn about lexical bundles and understand the meaning of the words which make up the bundles. They would be better acquainted with the communicative and discourse functions conveyed by those bundles in their disciplines. They then are more empowered as learners aided by language which could lead to better learning and understanding of their academic lectures.

As the findings of this research show, the use of lexical bundles and their functions varies across different academic divisions. Therefore, instructing students of different disciplines regarding the types and functions of bundles and how they are used to perform the functions related can be highly valuable. For example, the need to emphasize the use of directive bundles such as *have a look at, if you look at,* and *you can see that* or obligation bundles *we have to be, I want you to* and *so you have to* in performing laboratory tasks in Life and Medical Sciences.

It is widely realized that direct exposure to lexical bundles does not guarantee their acquisition and does not appear to be effective in raising the learners' awareness towards the existence of these lexical bundles in language. Rather it is considered necessary to

teach these frequently occurring bundles and put them into teaching materials which can then be designed to help students realize the structure and discourse functions of these bundles in academic lectures, especially those presented in the specific disciplinary division. In addition, the students' awareness needs to be raised towards the importance of lexical bundles in creating fluency in speaking. The findings of this study could to a large extent, help EAP textbook developers to utilize a list of the prevalent lexical bundles in the spoken language of different academic disciplines and how they are used to convey disciplinary functions.

Moreover, the findings can be truly beneficial for the field at large and in particular, novice lecturers in EFL/ ESL context, who are required to deliver the disciplinary materials in English. Efficient use of formulaic patterns such as lexical bundles in delivering academic disciplinary lectures provides opportunity for the lecturers' language to be more naturalistic and contain elements of native-like fluency. This would help them to communicate with members in their discourse community with an academic voice, and also to organize their discourse, since lexical bundles are one of the main sources of achieving coherence in language (Hyland, 2008). The study also helps in compiling an academic wordlist containing the commonly-used lexical bundles and the discourse functions they serve in academic lectures of a certain discipline. The list itself could be beneficial for raising the awareness of effective language use related to lexical bundles especially for novice lecturers and researchers as a source of reference (see Appendix B & D).

## REFERENCES


Ädel, A. (2008). What uh the folks who did this survey found: Expert attribution in spoken academic lectures. *Nordic Journal of English Studies*, *7*(3), 83-102.

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

AlHassan, L., & Wood, D. (2015). The effectiveness of focused instruction of formulaic sequences in augmenting L2 learners' academic writing skills: A quantitative research study. *Journal of English for Academic Purposes*, *17*, 51-62.

Allen, D. (2009). Lexical bundles in learner writing: An analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education, 1,* 105-127.

Altenberg, B. (1993). Recurrent word combinations in spoken English. In *Proceedings of the fifth Nordic association for English studies conference* (pp. 17-27). Reykjavik, Iceland: University of Iceland.

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

Anthony, L. (2014). AntConc (Version 3.4.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Bal, B. (2010). *Analysis of four-word lexical bundles in published research articles written by Turkish scholars* (Master's thesis, Georgia State University). Retrieved from https://scholarworks.gsu.edu/alesl_theses/2/

Becher, T. (1987). Disciplinary discourse. *Studies in Higher Education*, *12*(3), 261-274.

Becher, T. (1994). The significance of disciplinary differences. *Studies in Higher Education*, *19*(2), 151-161.

Beng, C. O. S., & Keong, Y. C. (2014). A corpus study of structural types of lexical bundles in MUET reading texts. *3L: Language, Linguistics, Literature®*, *20*(2), 127-140.

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

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

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers* (Vol. 23). Antwerp: John Benjamins Publishing.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, *26*(3), 263-286.

Butler, C. (1998). Collocational frameworks in Spanish. *International Journal of Corpus Linguistics, 3*(1), 1-32.

Bychkovska, T., & Lee, J. J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes*, *30*, 38-52.

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

Charles, M. (2003). 'This mystery…': a corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines. *Journal of English for Academic Purposes*, *2*(4), 313-326.

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

Chen, X., & Xiao, G. (2015, July). A Corpus-based Study of the Structures and Functions of Academic Chunks in the Discipline of Engineering. In *5th International Conference on Information Engineering for Mechanics and Materials*. Atlantis Press.

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

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: M.I.T. Press.

Conrad, S. M. (1996). Investigating academic texts with corpus-based techniques: An example from biology. *Linguistics and Education*, *8*(3), 299-326.

Conrad, S., & Biber, D. (2004). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica*, *20*, 56-71.

Cortes, V. (2002). *Lexical bundles in published and student academic writing in history and biology* (Doctoral dissertation, Northern Arizona University). Retrieved from https://www.researchgate.net/publication/222518326_Lexical_bundles_in_published_an

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

Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, *17*(4), 391-406.

Cortes, V. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora*, *3*(1), 43-57.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213–238.

Csomay, E. (2005). Linguistic variation in the lexical episodes of university classroom talk. *Language in use. Cognitive and discourse perspectives on language and language learning. Georgetown University round table on languages and linguistics*, 150-162.

Cortes, V., & Csomay, E. (2007). Positioning lexical bundles in university lectures. *Spoken Corpora in Applied Linguistics*, 55-77.

Csomay, E., & Cortes, V. (2010). Lexical bundle distribution in university classroom talk. *Language & Computers*, *71*(1), 153-168.

Csomay, E. (2013). Lexical bundles in discourse structure: A corpus-based study of classroom discourse. *Applied Linguistics*, *34*(3), 369-388.

Crompton, P. (1997). Hedging in academic writing: Some theoretical problems. *English for Specific Purposes*, *16*(4), 271-287.

DeCarrico, J., & Nattinger, J. R. (1988). Lexical phrases for the comprehension of academic lectures. *English for Specific Purposes*, *7*(2), 91-102.

De Cock, S. (1998). A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English. *International Journal of Corpus Linguistics, 3*(1), 59–80.

De Cock, S. (2003). *Recurrent sequences of words in native speaker and advanced learner spoken and written English: A corpus-driven approach* (Doctoral dissertation, Catholic University of Louvain). Retrieved from http://hdl.handle.net/2078.1/149775

De Cock, S. (2004). Preferred sequences of words in NS and NNS speech. *Belgian Journal of English Language and Literatures. BELL-Belgian Journal of English Language and Literature. 2*(2), 225-246.

Deroey, K. L., & Taverniers, M. (2012). Just remember this: Lexicogrammatical relevance markers in lectures. *English for Specific Purposes*, *31*(4), 221-233.

Dontcheva-Navratilova, O. (2012) 'Lexical bundles in academic texts by non-native speakers.' *Brno Studies in English*. *38* (2), 37-58.

Durrant, P., & Mathews-Aydınlı, J. (2011). A function-first approach to identifying formulaic language in academic writing. *English for Specific Purposes*, *30*(1), 58-72.

Durrant, P. (2013). Discipline and level specificity in university students' written vocabulary. *Applied Linguistics*, *35*(3), 328-356.

Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, *38*(2), 165-193.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text, 20*(1), 29–62.

Erman, B. (2009). Formulaic language from a learner perspective: What the learner needs to know. In R. Corrigan, E. A. Moravcssik, H. Ouali, and K.M. Wheatley (Eds.), *Formulaic Language*, Volume 2, (pp. 323-346) Philadelphia: John Benjamins Publishing Company.

Esfandiari, R., & Tavakoli Moein, G. (2016). A corpus-driven investigation into lexical bundles across research articles in Food Science and Technology. *English Language Teaching*, *3*(1), 30-1.

Farvardin, M. T., Afghari, A., & Koosha, M. (2012). Analysis of four-word lexical bundles in Physics research. *Advances in Digital Multimedia (ADMM). 1*(3), 2166-2916.

Ferguson, G. (2001). If you pop over there: A corpus-based study of conditionals in medical discourse. *English for Specific Purposes, 20*(1), 61–82.

Firth, J. R. (1957). *Papers in linguistics* 1934-1951. London: Oxford University Press.

Flowerdew, J. (Ed.). (2002). *Academic discourse.* New York, NY: Longman

Grabowski, L. (2013). Register variation across English pharmaceutical texts: A corpus-driven study of keywords, lexical bundles and phrase frames in patient information leaflets and summaries of product characteristics.' *Procedia – Social and Behavioral Sciences. 95*, 391-401.

Grabowski, L. (2015). Keywords and lexical bundles within English pharmaceutical discourse: A corpus-driven description. *English for Specific Purposes*, *38*, 23-33.

Greaves, C., & Warren, M. (2010). What can a corpus tell us about multi-word units. *The Routledge Handbook of Corpus Linguistics*, 212-226.

Güngör, F., (2016). *Crosslinguistic analysis of lexical bundles in L1 English, L2 English, and L1 Turkish research articles* (Doctoral dissertation, Gazi University). Retrieved from www.acikarsiv.gazi.edu.tr/File.php?Doc_ID=14643

Güngör, F., & Uysal, H. H. (2016). A comparative analysis of lexical bundles used by native and non-native scholars. *English Language Teaching*, *9*(6), 176-188.

Herbel-Eisenmann, B., & Wagner, D. (2010). Appraising lexical bundles in mathematics classroom discourse: Obligation and choice. *Educational Studies in Mathematics*, *75*(1), 43-63.

Herbel-Eisenmann, B., Wagner, D., & Cortes, V. (2010). Lexical bundle analysis in mathematics classroom discourse: The significance of stance. *Educational Studies in Mathematics*, *75*(1), 23-42.

Hernández, P. S. (2013). Lexical bundles in three oral corpora of university students. *Nordic Journal of English Studies*, *12*(1), 187-209.

Hewings, M. (Ed.). (2001). *Academic Writing in Context: Implications and Applications*. Birmingham: The University of Birmingham Press.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Abington, UK: Routledge.

Holmes, R. (1997). Genre analysis, and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes*, *16*(4), 321-337.

Huang, K. (2015). More does not mean better: Frequency and accuracy analysis of lexical bundles in Chinese EFL learners' essay writing. *System*, *53*, 13-23.

Huimin, S. (2014). *Moves and lexical bundles: A contrastive study of English agricultural science research articles between Chinese journals and internationally published journals* (Doctoral dissertation, Suranaree University of Technology). Retrieved from sutir.sut.ac.th:8080/sutir/bitstream/123456789/5342/2/Fulltext.Pdf

Hunston, S. (1995). A corpus study of some English verbs of attribution. *Functions of Language*, *2*(2), 133–158.

Hyland, K. (1994). Hedging in academic writing and EAF textbooks. *English for Specific Purposes*, *13*(3), 239-256.

Hyland, K. (1998). Persuasion and context: The pragmatics of academic metadiscourse. *Journal of Pragmatics*, *30*(4), 437-455.

Hyland, K. (2002). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics*, *34*(8), 1091-1112.

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"?. *TESOL Quarterly*, *41*(2), 235-253.

Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, *18*(1), 41-62.

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

Jablonkai, R. (2010). English in the context of European integration: A corpus-driven analysis of lexical bundles in English EU documents. *English for Specific Purposes*, *29* (4), 253–267.

Jalali, H., Rasekh, A. E., & Rizi, M. T. (2008). Lexical bundles and intradisciplinary variation: the case of applied linguistics. *Iranian Journal of Language Studies (IJLS)*, *2*(4), 447-484.

Jalali, H. (2009). *Lexical bundles in applied linguistics: variations within a single discipline*. (Unpublished doctoral dissertation). University of Isfahan, Isfahan, Iran.

Jalali, H. (2013). Lexical bundles in applied linguistics: Variations across postgraduate genres. *Journal of Foreign Language Teaching and Translation Studies*, *2*(2), 1-29.

Jalali, Z. S., Moini, M. R., & Arani, M. A. (2015). Structural and functional analysis of lexical bundles in medical research articles: A corpus-based study. *International Journal of Information*, *13*(1), 51-69.

Jalilifar, A., Ghoreishi, S. M., & Emam Roodband, S. A. (2016). Developing an inventory of core lexical bundles in English research articles: a cross-disciplinary corpus-based study. *Journal of World Languages*, *3*(3), 184-203.

Jespersen, O. (1924). *The philosophy of grammar*. London: Allen & Unwin.

Johnston, K. M. (2017). *Lexical Bundles in Applied Linguistics and Literature Writing: A Comparison of Intermediate English Learners and Professionals* (Doctoral dissertation, Portland State University). Retrieved from https://pdxscholar.library.pdx.edu/open_access_etds/3482/

Kaplan, R. B., & Grabe, W. (2002). A modern history of written discourse analysis. *Journal of Second Language Writing*, *11*(3), 191-223.

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

Kashiha, H., & Chan, S. H. (2013). An exploration of lexical bundles in academic lectures: examples from hard and soft sciences. *The Journal of Asia TEFL*, *10*(4), 133-161.

Kashiha, H., & Chan, S. H. (2014a). Structural analysis of lexical bundles in university lectures of politics and chemistry. *International Journal of Applied Linguistics and English Literature*, *3*(1), 224-230.

Kashiha, H., & Heng, C. S. (2014). Using multi-word units to take a stance in academic lectures. *Language & Communication*, *1*(1), 31-39.

Kashiha, H., & Chan, S. H. (2014b). Discourse functions of formulaic sequences in academic speech across two disciplines. *GEMA: Online Journal of Language Studies*, *14*(2), 15-27.

Kashiha, H., & Heng, C. S. (2014). Discourse functions of formulaic sequences in academic speech across two disciplines. *GEMA Online® Journal of Language Studies*, *14*(2), 15-27.

Kashiha, H. (2015). *Corpus-based study of lexical bundles in academic lectures across three disciplinary divisions*. (Unpublished doctoral dissertation). University Putra Malaysia, Malaysia.

Kim, C. K. (2009). Personal pronouns in English and Korean texts: A corpus-based study in terms of textual interaction. *Journal of Pragmatics*, *41*(10), 2086-2099.

Kuo, C. W. (2017). Hard or soft Science? Conceptualizing educational technology through a lexical bundle analytical approach. *Education and Linguistics Research*, *3*(1), 43-68.

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

Larsson Aas, H. (2011). *Recurrent word-combinations in spoken learner English: A study of corpus data from Swedish and Norwegian advanced learners*. Oslo: University of Oslo.

Lewis, M. (1997). *Implementing the lexical approach: Putting theory into practice.* Hove, England: Language Teaching Publications.

Li, L. (2016). *Sentence initial bundles in L2 thesis writing: A comparative study of Chinese L2 and New Zealand L1 postgraduates' writing* (Doctoral dissertation, University of Waikato). Retrieved from https://hdl.handle.net/10289/10862

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

Ma, G.H. (2009). A study of four-word lexical bundles in English major students' timed writing. *Foreign Language Teaching and Research, 41*(1), 54-60.

Marco, M. J. L. (2000). Collocational frameworks in medical research papers: a genre-based study. *English for Specific Purposes*, *19*(1), 63‑86.

Maswana, S., Kanamaru, T., & Tajino, A. (2013). Analyzing the journal corpus data on English expressions across disciplines. *The Journal of Asia TEFL*, *10*(4), 71-96.

Mbodj-Diop, N. B. (2016). *Lexical bundles in medical research articles: Structures and functions.* (Doctoral dissertation, Michigan State University). Available from ProQuest Dissertations & Theses database. (UMI No. 10107571)

McCarthy, M., & Carter, R. (2006). This that and the other: Multi-word clusters in spoken English as visible patterns of interaction. In M. McCarthy (Ed.), *Explorations in Corpus Linguistics* (pp. 7-26). New York: Cambridge University Press. (Reprinted from *Teanga: The Irish Yearbook of Applied Linguistics, 21*, 30-53, 2004).

Meunier, F., & Granger, S. (Eds.). (2008). *Phraseology in Foreign Language Learning and Teaching*. Amsterdam/Philadelphia: John Benjamins.

Moon, R., (1998). *Fixed Expressions and Idioms in English: A corpus-based approach*. Oxford: Oxford University Press.

Myers, G. (1989). The pragmatics of politeness in scientific articles. *Applied Linguistics*, *10*(1), 1-35.

Myers, G. (1990). The rhetoric of irony in academic writing. *Written Communication*, *7*(4), 419-455.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York, NY: Newbury House.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Neely, E., & Cortes, V. (2009). A little bit about: analyzing and teaching lexical bundles in academic lectures. *Language Value*, *1*(1), 17-38.

Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics*, *11*(3), 283-304.

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

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

Pang, W. (2010). Lexical bundles and the construction of an academic voice: A pedagogical perspective. *Asian EFL Journal*, *47*(1), 10-11.

Parvizi, N. (2011). *Identification of discipline-specific lexical bundles in education*. (Unpublished doctoral dissertation). University of Kashan, Isfahan, Iran.

Paquot, M. (2008). Exemplification in learner writing: A cross-linguistic perspective. In F. Meunier and S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 110-120). Amsterdam, PA: John Benjamins Publishing Company.

Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: native like selection and native like fluency. In J. Richards & R. Schmidt (Eds.), *Language and communication* (pp. 191-226). London: Longman.

Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes*, *14*, 84-94.

Pičuljan, M. (2016). *A Corpus-based Exploration of the Differences between Native and Non-native Texts in Academic English* (Doctoral dissertation, University of Rijeka). Retrieved from https://urn.nsk.hr/urn:nbn:hr:186:934036

Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System*, *42*, 220-231.

Rafiee, M., Tavakoli, M., & Amirian, Z. (2011). Structural analysis of lexical bundles across two types of English newspapers edited by native and non-native speakers. *Modern Journal of Applied Linguistics*, *3*(2), 218-236.

Reppen, R. (2004). Academic language: An exploration of university classroom and textbook language. *Discourse in the Professions: Perspectives from Corpus Linguistics* (pp. 65-86). Amsterdam: John Benjamins Publishing Company.

Salager-Meyer, F. (1994). Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes, 13*(2), 149-171.

Salazar, D. J. L. (2011). *Lexical bundles in scientific English: A corpus-based study of native and non-native writing* (Doctoral dissertation, University of Barcelona). Retrieved from http://hdl.handle.net/2445/35023

Salazar, D. J. L. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching* (Vol. 65). Amsterdam: John Benjamins Publishing Company.

Schleef, E. (2008). The "lecturer's OK" revisited: changing discourse conventions and the influence of academic division. *American Speech*, *83*(1), 62-84.

Schmitt, N. & McCarthy, M. (Eds.). (1997). *Vocabulary: Description, acquisition and pedagogy*. Cambridge: Cambridge University Press.

Schmitt, N., & Carter, R. (2004). *Formulaic sequences in action. Formulaic Sequences: Acquisition, processing and use* (Vol. 9). Amsterdam: John Benjamins Publishing.

Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education* (Vol. 22). Amsterdam: John Benjamins Publishing.

Shamsudin, S., Yusof, M. A. M., & Raof, A. H. A. (2012). Lecture language in Malaysian engineering lecture corpus. *Procedia-Social and Behavioral Sciences*, *56*, 298-306.

Silver, M. (2003). The stance of stance: a critical look at ways stance is expressed and modeled in academic discourse. *Journal of English for Academic Purposes*, *2*(4), 359-374.

Simpson, R. C., & Swales, J. M. (Eds.). (2001). *Corpus linguistics in North America: Selections from the 1999 symposium*. Michigan: University of Michigan Press.

Simpson, R. (2004). Stylistic features of academic speech: The role of formulaic expressions. *Discourse in the professions: Perspectives from Corpus Linguistics*, 37-64.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, *31*(4), 487-512.

Sinclair, J. M. (Ed.). (1987). *Looking up: An account of the COBUILD project in lexical computing and the development of the Collins COBUILD English language dictionary*. London: HarperCollins.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, *12*(3), 214-225.

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

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Thompson, N. (2003). *Communication and language*. New York: Palgrave MacMillan.

Tománková, V. (2016). Lexical bundles in legal texts corpora–selection, classification and pedagogical implications. *Discourse and Interaction*, *9*(2), 75-94.

Tracy-Ventura, N., Cortes, V., & Biber, D. (2007). Lexical bundles in speech and writing. In G. Parodi (Ed.), *Working with Spanish Corpora* (pp. 217-230). London, UK: Continuum.

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

Varttala, T. (2003). Hedging in scientific research articles: A cross-disciplinary study. In G. Cortese & P. Riley (Eds.), *Domain-Specific English: Textual Practices across Communities and Classrooms* (pp. 141–174). New York, NY: Peter Lang.

Vidakovic, I., & Barker, F. (2010). Use of words and multi-word units in skills for life writing examinations. *Cambridge ESOL: Research Notes*, *41*, 7-14.

Wang, Y. (2017). Lexical bundles in spoken academic ELF. *International Journal of Corpus Linguistics*, *22*(2), 187-211.

Wang, Y. (2018). Formulaic sequences signalling discourse organisation in ELF academic lectures: a disciplinary perspective. *Journal of English as a Lingua Franca*, *7*(2), 355-376.

Wei, N.X. (2007). Phraseological characteristics of Chinese learners' spoken English: Evidence of lexical chunks from COLSEC. *Modern Foreign Languages, 30*, 281-291.

Weinstein, M. (2011). *Investigating the lexical bundle use of novice and upper-level writers within a university discipline* (Doctoral dissertation, Carleton University). Available from ProQuest Dissertation & Theses database. (UMI No. MR83101)

Willis, D. (1990). *The lexical syllabus*. London: Collins.

Wood, D. C., & Appel, R. (2014). Multiword constructions in first year business and engineering university textbooks and EAP textbooks. *Journal of English for Academic Purposes*, *15*, 1-13.

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

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

Wray, A. (2002). Formulaic language in computer-supported communication: theory meets reality. *Language Awareness*, *11*(2), 114-131.

Xu, J., & Xu, Z. (2007). Discourse management chunks in Chinese college learners' English speech: A spoken corpus-based study. *Foreign Language Teaching and Research, 6*(8), 437-443.

# APPENDICES

**APPENDIX A: Lexical bundles which were excluded from the Arts and Humanities sub-corpus:**

| No | Lexical bundles |
|---|---|
| **1** | of the nineteenth century |
| **2** | in the nineteenth century |
| **3** | in the twentieth century |
| **4** | in the in the |
| **5** | the the the the |
| **6** | of the of the |
| **7** | in a in a |
| **8** | in the eighteenth century |
| **9** | of the twentieth century |
| **10** | the Second World War |
| **11** | on the on the |
| **12** | er sort of er |
| **13** | sort of er er |
| **14** | as a as a |
| **15** | er er er er |
| **16** | er er sort of |
| **17** | er one of the |
| **18** | er the the the |
| **19** | of the eighteenth century |
| **20** | er some of the |
| **21** | the late nineteenth century |
| **22** | the nineteenth century and |
| **23** | a kind of er |
| **24** | if you like er |
| **25** | the er er er |
| **26** | the er the the |
| **27** | the First World War |
| **28** | of er er er |
| **29** | the the sort of |
| **30** | a a a a |
| **31** | a sort of er |
| **32** | er a kind of |
| **33** | er a sort of |
| **34** | in the er er |
| **35** | of the er of |
| **36** | a lot of er |
| **37** | er er in the |
| **38** | er the er the |
| **39** | in in in in |
| **40** | in the er in |
| **41** | in the late nineteenth |
| **42** | the er in the |
| **43** | er at the time |
| **44** | er in other words |

| | |
|---|---|
| **45** | er in the er |
| **46** | er the way in |
| **47** | it's a it's a |
| **48** | of the er the |
| **49** | some of the er |
| **50** | the development of capitalism |
| **51** | the er er the |
| **52** | the er of the |
| **53** | the er you know |
| **54** | this is a a |
| **55** | you know the the |
| **56** | er and i think |
| **57** | er and of course |
| **58** | er er the the |
| **59** | er in a sense |
| **60** | er in a way |
| **61** | er the sort of |
| **62** | er this is a |
| **63** | er you know the |
| **64** | if you if you |
| **65** | in the age of |
| **66** | in the er the |
| **67** | in the in in |
| **68** | in the sixteenth century |
| **69** | of a of a |
| **70** | of the late nineteenth |
| **71** | one of the er |
| **72** | the development of the |
| **73** | the eighteenth century er |
| **74** | the the er the |
| **75** | the twentieth century and |
| **76** | their own history but |
| **77** | this er sort of |
| **78** | this sort of er |
| **79** | way in which history |

**APPENDIX B: Target bundles in Arts and Humanities sub-corpus:**

| No | Lexical bundles | Frequency |
|---|---|---|
| **1** | the way in which | 77(12.1%) |
| **2** | the end of the | 60(9.4%) |
| **3** | at the end of | 49(7.7%) |
| **4** | if you look at | 37(5.8%) |
| **5** | on the other hand | 31(4.8%) |
| **6** | what i want to | 31(4.8%) |
| **7** | at the same time | 30(4.7%) |
| **8** | in other words the | 30(9.7%) |
| **9** | is one of the | 30(9.7%) |
| **10** | on the one hand | 29(9.3%) |
| **11** | and so on and | 28(9%) |
| **12** | one of the things | 28(9%) |
| **13** | in terms of the | 26(8.4%) |
| **14** | and one of the | 25(8%) |
| **15** | the idea of the | 24(7.7%) |
| **16** | all the rest of | 23(7.4%) |
| **17** | as a kind of | 23(7.4%) |
| **18** | the rest of it | 23(7.4%) |
| **19** | to be able to | 22(7.1%) |
| **20** | and all the rest | 21(6.7%) |
| **21** | at the beginning of | 21(6.7%) |
| **22** | the rest of the | 21(6.7%) |
| **23** | and you can see | 19(6.1%) |
| **24** | so in other words | 19(6.1%) |
| **25** | to do with the | 19(6.1%) |
| **26** | a lot of the | 18(5.8%) |
| **27** | and this is the | 18(5.8%) |
| **28** | i don't know if | 18(5.8%) |
| **29** | in the context of | 18(5.8%) |
| **30** | you look at the | 18(5.8%) |
| **31** | i want to do | 17(5.5%) |
| **32** | if you like the | 17(5.5%) |
| **33** | in the way that | 17(5.5%) |
| **34** | so on and so | 17(5.5%) |
| **35** | and i'm going to | 16(5.1%) |
| **36** | as a sort of | 16(5.1%) |
| **37** | is a kind of | 16(5.1%) |
| **38** | what i'm going to | 16(5.1%) |
| **39** | you know this is | 16(5.1%) |
| **40** | if you like of | 15(4.8%) |
| **41** | if you want to | 15(4.8%) |
| **42** | on and so forth | 15(4.8%) |
| **43** | one of the most | 15(4.8%) |
| **44** | the nature of the | 15(4.8%) |
| **45** | and this is a | 14(4.5%) |
| **46** | at the heart of | 14(4.5%) |
| **47** | i'm not going to | 14(4.5%) |

| | | |
|---|---|---|
| **48** | is going to be | 14(4.5%) |
| **49** | one of the reasons | 14(4.5%) |
| **50** | the beginning of the | 14(4.5%) |
| **51** | the idea of a | 14(4.5%) |
| **52** | those of you who | 14(4.5%) |
| **53** | way in which the | 14(4.5%) |
| **54** | you can see that | 14(4.5%) |
| **55** | are going to be | 13(4.2%) |
| **56** | the way that the | 13(4.2%) |
| **57** | a little bit more | 12(3.8%) |
| **58** | a whole series of | 12(3.8%) |
| **59** | as you can see | 12(3.8%) |
| **60** | at the centre of | 12(3.8%) |
| **61** | at the start of | 12(3.8%) |
| **62** | is the kind of | 12(3.8%) |
| **63** | sort of you know | 12(3.8%) |
| **64** | the meaning of the | 12(3.8%) |
| **65** | the ways in which | 12(3.8%) |
| **66** | there's a kind of | 12(3.8%) |
| **67** | this is one of | 12(3.8%) |
| **68** | to do is to | 12(3.8%) |
| **69** | to look at the | 12(3.8%) |
| **70** | which i think is | 12(3.8%) |
| **71** | i just want to | 11(3.5%) |
| **72** | in so far as | 11(3.5%) |
| **73** | in the case of | 11(3.5%) |
| **74** | one of the great | 11(3.5%) |
| **75** | or something like that | 11(3.5%) |
| **76** | point of view of | 11(3.5%) |
| **77** | some of you may | 11(3.5%) |
| **78** | that there is a | 11(3.5%) |
| **79** | the interests of the | 11(3.5%) |
| **80** | the middle of the | 11(3.5%) |
| **81** | the point of view | 11(3.5%) |
| **82** | the whole of the | 11(3.5%) |
| **83** | was a kind of | 11(3.5%) |
| **84** | a little bit about | 10(3.2%) |
| **85** | a little bit of | 10(3.2%) |
| **86** | an idea of a | 10(3.2%) |
| **87** | at the level of | 10(3.2%) |
| **88** | for the first time | 10(3.2%) |
| **89** | going to be the | 10(3.2%) |
| **90** | going to look at | 10(3.2%) |
| **91** | going to show you | 10(3.2%) |
| **92** | going to talk about | 10(3.2%) |
| **93** | i'll come back to | 10(3.2%) |
| **94** | i'm going to show | 10(3.2%) |
| **95** | i'm just going to | 10(3.2%) |
| **96** | if you like in | 10(3.2%) |
| **97** | in a way that | 10(3.2%) |

| | | |
|---|---|---|
| **98** | in front of you | 10(3.2%) |
| **99** | in terms of what | 10(3.2%) |
| **100** | in the first place | 10(3.2%) |
| **101** | in the same way | 10(3.2%) |
| **102** | is a sort of | 10(3.2%) |
| **103** | it has to be | 10(3.2%) |
| **104** | of course in the | 10(3.2%) |
| **105** | of the sort of | 10(3.2%) |
| **106** | of the things that | 10(3.2%) |
| **107** | the course of the | 10(3.2%) |
| **108** | the start of the | 10(3.2%) |
| **109** | the turn of the | 10(3.2%) |
| **110** | to the idea of | 10(3.2%) |
| **111** | was one of the | 10(3.2%) |
| **112** | way in which we | 10(3.2%) |
| **113** | a kind of a | 9(2.9%) |
| **114** | about the nature of | 9(2.9%) |
| **115** | and of course the | 9(2.9%) |
| **116** | at the top of | 9(2.9%) |
| **117** | draw your attention to | 9(2.9%) |
| **118** | from the point of | 9(2.9%) |
| **119** | going to be a | 9(2.9%) |
| **120** | i don't want to | 9(2.9%) |
| **121** | i think it's a | 9(2.9%) |
| **122** | i'm going to be | 9(2.9%) |
| **123** | i'm going to do | 9(2.9%) |
| **124** | if you like to | 9(2.9%) |
| **125** | in a number of | 9(2.9%) |
| **126** | in a sort of | 9(2.9%) |
| **127** | in terms of a | 9(2.9%) |
| **128** | in the middle of | 9(2.9%) |
| **129** | is the way in | 9(2.9%) |
| **130** | it seems to me | 9(2.9%) |
| **131** | it's a sort of | 9(2.9%) |
| **132** | of the kind of | 9(2.9%) |
| **133** | so i'm going to | 9(2.9%) |
| **134** | so that you can | 9(2.9%) |
| **135** | the idea that the | 9(2.9%) |
| **136** | to give you a | 9(2.9%) |
| **137** | what it is to | 9(2.9%) |
| **138** | what it means to | 9(2.9%) |
| **139** | what we're going to | 9(2.9%) |
| **140** | about the way in | 8(2.5%) |
| **141** | and as i say | 8(2.5%) |
| **142** | and the idea of | 8(2.5%) |
| **143** | and the way in | 8(2.5%) |
| **144** | as a result of | 8(2.5%) |
| **145** | but i think it's | 8(2.5%) |
| **146** | by the end of | 8(2.5%) |
| **147** | for those of you | 8(2.5%) |

| | | |
|---|---|---|
| **148** | have a look at | 8(2.5%) |
| **149** | i mean this is | 8(2.5%) |
| **150** | i want you to | 8(2.5%) |
| **151** | if you think of | 8(2.5%) |
| **152** | in a kind of | 8(2.5%) |
| **153** | in a way which | 8(2.5%) |
| **154** | in relation to the | 8(2.5%) |
| **155** | in this sort of | 8(2.5%) |
| **156** | is part of the | 8(2.5%) |
| **157** | it's going to be | 8(2.5%) |
| **158** | of the reasons why | 8(2.5%) |
| **159** | that i want to | 8(2.5%) |
| **160** | that is to say | 8(2.5%) |
| **161** | that kind of thing | 8(2.5%) |
| **162** | the back of the | 8(2.5%) |
| **163** | the fact that the | 8(2.5%) |
| **164** | the one of the | 8(2.5%) |
| **165** | the second half of | 8(2.5%) |
| **166** | to a certain extent | 8(2.5%) |
| **167** | to bear in mind | 8(2.5%) |
| **168** | to come back to | 8(2.5%) |
| **169** | to think about the | 8(2.5%) |
| **170** | towards the end of | 8(2.5%) |
| **171** | we're going to be | 8(2.5%) |
| **172** | what's going on in | 8(2.5%) |
| **173** | when we look at | 8(2.5%) |
| **174** | you know in the | 8(2.5%) |
| **175** | you might want to | 8(2.5%) |
| **176** | a great deal of | 7(2.2%) |
| **177** | a way in which | 7(2.2%) |
| **178** | and i want to | 7(2.2%) |
| **179** | and on the other | 7(2.2%) |
| **180** | and so on but | 7(2.2%) |
| **181** | and that is the | 7(2.2%) |
| **182** | and the fact that | 7(2.2%) |
| **183** | as well as the | 7(2.2%) |
| **184** | come back to that | 7(2.2%) |
| **185** | come back to this | 7(2.2%) |
| **186** | going on in the | 7(2.2%) |
| **187** | have to look at | 7(2.2%) |
| **188** | i think this is | 7(2.2%) |
| **189** | if we look at | 7(2.2%) |
| **190** | if you like and | 7(2.2%) |
| **191** | in a sense the | 7(2.2%) |
| **192** | in such a way | 7(2.2%) |
| **193** | in the course of | 7(2.2%) |
| **194** | in the sense that | 7(2.2%) |
| **195** | in the sort of | 7(2.2%) |
| **196** | is in fact a | 7(2.2%) |
| **197** | is the idea that | 7(2.2%) |

| | | |
|---|---|---|
| **198** | it in terms of | 7(2.2%) |
| **199** | it may be that | 7(2.2%) |
| **200** | it's not just that | 7(2.2%) |
| **201** | just to remind you | 7(2.2%) |
| **202** | not going to be | 7(2.2%) |
| **203** | now i want to | 7(2.2%) |
| **204** | of the things which | 7(2.2%) |
| **205** | of the ways in | 7(2.2%) |
| **206** | of you who are | 7(2.2%) |
| **207** | second half of the | 7(2.2%) |
| **208** | so this is the | 7(2.2%) |
| **209** | so you can see | 7(2.2%) |
| **210** | something to do with | 7(2.2%) |
| **211** | that in a minute | 7(2.2%) |
| **212** | that's one of the | 7(2.2%) |
| **213** | the fact that he | 7(2.2%) |
| **214** | the fact that it | 7(2.2%) |
| **215** | the first part of | 7(2.2%) |
| **216** | the heart of the | 7(2.2%) |
| **217** | the people who are | 7(2.2%) |
| **218** | the same kind of | 7(2.2%) |
| **219** | this is the kind | 7(2.2%) |
| **220** | to say that the | 7(2.2%) |
| **221** | to the fact that | 7(2.2%) |
| **222** | want to do is | 7(2.2%) |
| **223** | was going to be | 7(2.2%) |
| **224** | you can see the | 7(2.2%) |
| **225** | you have to remember | 7(2.2%) |
| **226** | you know if you | 7(2.2%) |
| **227** | you look at it | 7(2.2%) |
| **228** | you're going to be | 7(2.2%) |
| **229** | a bit of a | 6(1.9%) |
| **230** | a certain amount of | 6(1.9%) |
| **231** | a sort of a | 6(1.9%) |
| **232** | all sorts of things | 6(1.9%) |
| **233** | and he had a | 6(1.9%) |
| **234** | and i think the | 6(1.9%) |
| **235** | and i think this | 6(1.9%) |
| **236** | and that is a | 6(1.9%) |
| **237** | and that is that | 6(1.9%) |
| **238** | and think about the | 6(1.9%) |
| **239** | and this is what | 6(1.9%) |
| **240** | are we going to | 6(1.9%) |
| **241** | as it were a | 6(1.9%) |
| **242** | as well and the | 6(1.9%) |
| **243** | at some of the | 6(1.9%) |
| **244** | at the bottom of | 6(1.9%) |
| **245** | at the time and | 6(1.9%) |
| **246** | but i think the | 6(1.9%) |
| **247** | but on the other | 6(1.9%) |

| | | |
|---|---|---|
| **248** | but the point is | 6(1.9%) |
| **249** | can be seen as | 6(1.9%) |
| **250** | first of all the | 6(1.9%) |
| **251** | for the rest of | 6(1.9%) |
| **252** | give you an example | 6(1.9%) |
| **253** | i just wanted to | 6(1.9%) |
| **254** | i said at the | 6(1.9%) |
| **255** | i think you know | 6(1.9%) |
| **256** | i want to talk | 6(1.9%) |
| **257** | i was going to | 6(1.9%) |
| **258** | i was talking about | 6(1.9%) |
| **259** | if you like which | 6(1.9%) |
| **260** | if you're going to | 6(1.9%) |
| **261** | in one of his | 6(1.9%) |
| **262** | in which we think | 6(1.9%) |
| **263** | is to do with | 6(1.9%) |
| **264** | kind of you know | 6(1.9%) |
| **265** | know this is a | 6(1.9%) |
| **266** | look at some of | 6(1.9%) |
| **267** | now what i want | 6(1.9%) |
| **268** | of the fact that | 6(1.9%) |
| **269** | of this sort of | 6(1.9%) |
| **270** | of what's going on | 6(1.9%) |
| **271** | of you know the | 6(1.9%) |
| **272** | on the part of | 6(1.9%) |
| **273** | on the top of | 6(1.9%) |
| **274** | one way or another | 6(1.9%) |
| **275** | seems to me that | 6(1.9%) |
| **276** | so far as it | 6(1.9%) |
| **277** | so there is a | 6(1.9%) |
| **278** | so this is a | 6(1.9%) |
| **279** | that a lot of | 6(1.9%) |
| **280** | that there was a | 6(1.9%) |
| **281** | the bottom of the | 6(1.9%) |
| **282** | the centre of the | 6(1.9%) |
| **283** | the kind of the | 6(1.9%) |
| **284** | the one hand and | 6(1.9%) |
| **285** | the role of the | 6(1.9%) |
| **286** | the second part of | 6(1.9%) |
| **287** | the top of the | 6(1.9%) |
| **288** | the way that it | 6(1.9%) |
| **289** | the you know the | 6(1.9%) |
| **290** | there will be a | 6(1.9%) |
| **291** | there would be a | 6(1.9%) |
| **292** | there's a lot of | 6(1.9%) |
| **293** | there's a sort of | 6(1.9%) |
| **294** | there's going to be | 6(1.9%) |
| **295** | this idea of the | 6(1.9%) |
| **296** | this is something that | 6(1.9%) |
| **297** | this is what the | 6(1.9%) |

| | | |
|---|---|---|
| **298** | to be looking at | 6(1.9%) |
| **299** | to do now is | 6(1.9%) |
| **300** | to look at this | 6(1.9%) |
| **301** | to look at what | 6(1.9%) |
| **302** | to talk about the | 6(1.9%) |
| **303** | trying to do is | 6(1.9%) |
| **304** | want to think about | 6(1.9%) |
| **305** | we look at the | 6(1.9%) |
| **306** | we'll be looking at | 6(1.9%) |
| **307** | we're going to look | 6(1.9%) |
| **308** | what does it mean | 6(1.9%) |
| **309** | you like of the | 6(1.9%) |

**APPENDIX C: Lexical bundles which were excluded from the Life and Medical Sciences sub-corpus:**

| No | Lexical bundles |
|---|---|
| 1 | fifty per cent of |
| 2 | in the in the |
| 3 | in the u-k er |
| 4 | the the the the |
| 5 | twenty per cent of |
| 6 | er and this is |
| 7 | er one of the |
| 8 | er this is the |
| 9 | er and these are |
| 10 | er in terms of |
| 11 | er this is a |
| 12 | ten per cent of |
| 13 | et cetera et cetera |
| 14 | five per cent of |
| 15 | if you if you |
| 16 | at the moment er |
| 17 | eighty per cent of |
| 18 | er i don't know |
| 19 | in the u-k and |
| 20 | this is this is |
| 21 | going to be er |
| 22 | in a in a |
| 23 | of the of the |
| 24 | one two three four |
| 25 | protein in the urine |
| 26 | two three four five |
| 27 | a little bit er |
| 28 | er first of all |
| 29 | er i'm going to |
| 30 | er some of the |
| 31 | forty per cent of |
| 32 | in the u-k we |
| 33 | ninety per cent of |
| 34 | seventy per cent of |
| 35 | the clinical methods course |
| 36 | there are there are |
| 37 | this is a a |
| 38 | to that in a |
| 39 | er and there is |
| 40 | er but i think |
| 41 | er er er er |
| 42 | er if you have |
| 43 | er the the the |
| 44 | four five six seven |
| 45 | is a is a |
| 46 | it is it is |
| 47 | sixty per cent of |

| | |
|---|---|
| **48** | the development of the |
| **49** | this is er a |
| **50** | three four five six |

**APPENDIX D: Target bundles in the Life and Medical Sciences sub-corpus:**

| No | Lexical bundles | Frequency |
|---|---|---|
| **1** | going to talk about | 48(14.6%) |
| **2** | to be able to | 47(14.3%) |
| **3** | if you want to | 45(13.7%) |
| **4** | per cent of the | 45(13.7%) |
| **5** | you can see that | 41(12.5%) |
| **6** | i'm not going to | 40(12.2%) |
| **7** | at the end of | 39(11.9%) |
| **8** | in the case of | 36(11%) |
| **9** | the end of the | 36(11%) |
| **10** | and you can see | 35(10.7%) |
| **11** | is going to be | 33(10%) |
| **12** | i'm going to talk | 30(9.1%) |
| **13** | a little bit of | 28(8.5%) |
| **14** | are going to be | 28(8.5%) |
| **15** | if we look at | 26(7.9%) |
| **16** | one of the things | 26(7.9%) |
| **17** | and this is the | 25(7.6%) |
| **18** | going to look at | 25(7.6%) |
| **19** | if you look at | 24(7.3%) |
| **20** | and things like that | 23(7%) |
| **21** | of the things that | 23(7%) |
| **22** | and i'm going to | 22(6.7%) |
| **23** | a little bit about | 21(6.4%) |
| **24** | it's going to be | 20(6.1%) |
| **25** | thank you very much | 20(6.1%) |
| **26** | and this is a | 19(5.8%) |
| **27** | i don't know if | 19(5.8%) |
| **28** | per cent of people | 19(5.8%) |
| **29** | what i'm going to | 19(5.8%) |
| **30** | we're going to look | 18(5.5%) |
| **31** | at the same time | 17(5.1%) |
| **32** | i just want to | 17(5.1%) |
| **33** | if you have a | 16(4.8%) |
| **34** | quite a lot of | 16(4.8%) |
| **35** | so i'm going to | 16(4.8%) |
| **36** | so this is a | 16(4.8%) |
| **37** | a lot of the | 15(4.5%) |
| **38** | i want you to | 15(4.5%) |
| **39** | in terms of the | 15(4.5%) |
| **40** | so you have to | 15(4.5%) |
| **41** | there's a lot of | 15(4.5%) |
| **42** | those of you who | 15(4.5%) |
| **43** | what do you think | 15(4.5%) |
| **44** | what i want to | 15(4.5%) |
| **45** | you can see here | 15(4.5%) |

| | | |
|---|---|---|
| **46** | you look at the | 15(4.5%) |
| **47** | you need to know | 15(4.5%) |
| **48** | you're going to be | 15(4.5%) |
| **49** | can see that the | 14(4.2%) |
| **50** | going to be a | 14(4.2%) |
| **51** | if you think about | 14(4.2%) |
| **52** | not going to go | 14(4.2%) |
| **53** | so this is the | 14(4.2%) |
| **54** | the rest of the | 14(4.2%) |
| **55** | to do is to | 14(4.2%) |
| **56** | you can see the | 14(4.2%) |
| **57** | and one of the | 13(3.9%) |
| **58** | and this is what | 13(3.9%) |
| **59** | as you can see | 13(3.9%) |
| **60** | but you can see | 13(3.9%) |
| **61** | okay so this is | 13(3.9%) |
| **62** | on the other hand | 13(3.9%) |
| **63** | or something like that | 13(3.9%) |
| **64** | so you can see | 13(3.9%) |
| **65** | to get rid of | 13(3.9%) |
| **66** | to look at the | 13(3.9%) |
| **67** | you don't need to | 13(3.9%) |
| **68** | you think about it | 13(3.9%) |
| **69** | you're not going to | 13(3.9%) |
| **70** | a lot of people | 12(3.6%) |
| **71** | and if you look | 12(3.6%) |
| **72** | and we're going to | 12(3.6%) |
| **73** | go back to the | 12(3.6%) |
| **74** | going to go into | 12(3.6%) |
| **75** | is one of the | 12(3.6%) |
| **76** | the rest of it | 12(3.6%) |
| **77** | this part of the | 12(3.6%) |
| **78** | to talk about the | 12(3.6%) |
| **79** | we look at the | 12(3.6%) |
| **80** | we're going to be | 12(3.6%) |
| **81** | why do you think | 12(3.6%) |
| **82** | are you going to | 11(3.3%) |
| **83** | come back to that | 11(3.3%) |
| **84** | for a long time | 11(3.3%) |
| **85** | going to be the | 11(3.3%) |
| **86** | i'm going to do | 11(3.3%) |
| **87** | in the middle of | 11(3.3%) |
| **88** | not going to be | 11(3.3%) |
| **89** | over a period of | 11(3.3%) |
| **90** | so if we look | 11(3.3%) |
| **91** | the nature of the | 11(3.3%) |
| **92** | to bear in mind | 11(3.3%) |
| **93** | to do with the | 11(3.3%) |
| **94** | we're going to talk | 11(3.3%) |
| **95** | when we look at | 11(3.3%) |

| | | |
|---|---|---|
| **96** | you can see is | 11(3.3%) |
| **97** | you have to have | 11(3.3%) |
| **98** | a bit more about | 10(3%) |
| **99** | a little bit more | 10(3%) |
| **100** | and of course the | 10(3%) |
| **101** | and so on and | 10(3%) |
| **102** | and these are the | 10(3%) |
| **103** | and you have to | 10(3%) |
| **104** | are there any questions | 10(3%) |
| **105** | for those of you | 10(3%) |
| **106** | if you don't know | 10(3%) |
| **107** | in other words the | 10(3%) |
| **108** | in the context of | 10(3%) |
| **109** | just to remind you | 10(3%) |
| **110** | of you who are | 10(3%) |
| **111** | talk a little bit | 10(3%) |
| **112** | the way in which | 10(3%) |
| **113** | what i mean by | 10(3%) |
| **114** | what's going to happen | 10(3%) |
| **115** | a bit of a | 9(2.7%) |
| **116** | and this is because | 9(2.7%) |
| **117** | are a lot of | 9(2.7%) |
| **118** | at the back there | 9(2.7%) |
| **119** | bear that in mind | 9(2.7%) |
| **120** | if you've got a | 9(2.7%) |
| **121** | in the absence of | 9(2.7%) |
| **122** | it's one of the | 9(2.7%) |
| **123** | it's very difficult to | 9(2.7%) |
| **124** | nothing to do with | 9(2.7%) |
| **125** | one of the most | 9(2.7%) |
| **126** | one of the problems | 9(2.7%) |
| **127** | so if you have | 9(2.7%) |
| **128** | so these are the | 9(2.7%) |
| **129** | so we need to | 9(2.7%) |
| **130** | that you're going to | 9(2.7%) |
| **131** | there are a lot | 9(2.7%) |
| **132** | there are lots of | 9(2.7%) |
| **133** | they're going to be | 9(2.7%) |
| **134** | this is one of | 9(2.7%) |
| **135** | to give you a | 9(2.7%) |
| **136** | want to talk about | 9(2.7%) |
| **137** | we're not going to | 9(2.7%) |
| **138** | what we're going to | 9(2.7%) |
| **139** | you've got to have | 9(2.7%) |
| **140** | and a lot of | 8(2.4%) |
| **141** | and i don't know | 8(2.4%) |
| **142** | as we go along | 8(2.4%) |
| **143** | at the beginning of | 8(2.4%) |
| **144** | at the moment and | 8(2.4%) |
| **145** | but the point is | 8(2.4%) |

| | | |
|---|---|---|
| **146** | by the end of | 8(2.4%) |
| **147** | do you think it's | 8(2.4%) |
| **148** | i don't know how | 8(2.4%) |
| **149** | i don't know what | 8(2.4%) |
| **150** | i don't think it's | 8(2.4%) |
| **151** | i don't want to | 8(2.4%) |
| **152** | i said to you | 8(2.4%) |
| **153** | i want to do | 8(2.4%) |
| **154** | in any great detail | 8(2.4%) |
| **155** | in the same way | 8(2.4%) |
| **156** | is that if you | 8(2.4%) |
| **157** | not going to talk | 8(2.4%) |
| **158** | one of the reasons | 8(2.4%) |
| **159** | per cent of your | 8(2.4%) |
| **160** | so if we have | 8(2.4%) |
| **161** | so if you want | 8(2.4%) |
| **162** | so in other words | 8(2.4%) |
| **163** | so that you can | 8(2.4%) |
| **164** | so what i'm going | 8(2.4%) |
| **165** | that one of the | 8(2.4%) |
| **166** | the first thing that | 8(2.4%) |
| **167** | the most important thing | 8(2.4%) |
| **168** | the size of the | 8(2.4%) |
| **169** | there's going to be | 8(2.4%) |
| **170** | those of you that | 8(2.4%) |
| **171** | to make sure that | 8(2.4%) |
| **172** | to talk to you | 8(2.4%) |
| **173** | want to do is | 8(2.4%) |
| **174** | we have to be | 8(2.4%) |
| **175** | when it comes to | 8(2.4%) |
| **176** | you don't have to | 8(2.4%) |
| **177** | you end up with | 8(2.4%) |
| **178** | you're going to get | 8(2.4%) |
| **179** | a bit more detail | 7(2.1%) |
| **180** | a long period of | 7(2.1%) |
| **181** | a small amount of | 7(2.1%) |
| **182** | and i think the | 7(2.1%) |
| **183** | and if you do | 7(2.1%) |
| **184** | and if you don't | 7(2.1%) |
| **185** | and if you have | 7(2.1%) |
| **186** | and if you think | 7(2.1%) |
| **187** | and that is the | 7(2.1%) |
| **188** | and there is a | 7(2.1%) |
| **189** | as well as the | 7(2.1%) |
| **190** | be a little bit | 7(2.1%) |
| **191** | by the time you | 7(2.1%) |
| **192** | come and see me | 7(2.1%) |
| **193** | do you think the | 7(2.1%) |
| **194** | does that make sense | 7(2.1%) |
| **195** | get rid of the | 7(2.1%) |

| | | |
|---|---|---|
| **196** | going to show you | 7(2.1%) |
| **197** | have a look at | 7(2.1%) |
| **198** | i want to talk | 7(2.1%) |
| **199** | i was going to | 7(2.1%) |
| **200** | i'll come back to | 7(2.1%) |
| **201** | if we're going to | 7(2.1%) |
| **202** | if you do a | 7(2.1%) |
| **203** | if you forget everything | 7(2.1%) |
| **204** | if you're going to | 7(2.1%) |
| **205** | in a bit more | 7(2.1%) |
| **206** | in other words you | 7(2.1%) |
| **207** | is a little bit | 7(2.1%) |
| **208** | is the fact that | 7(2.1%) |
| **209** | is the sort of | 7(2.1%) |
| **210** | it has to be | 7(2.1%) |
| **211** | it used to be | 7(2.1%) |
| **212** | little bit about the | 7(2.1%) |
| **213** | most of the time | 7(2.1%) |
| **214** | now i'm going to | 7(2.1%) |
| **215** | so here we have | 7(2.1%) |
| **216** | so there is a | 7(2.1%) |
| **217** | so you need to | 7(2.1%) |
| **218** | so you've got a | 7(2.1%) |
| **219** | that i want to | 7(2.1%) |
| **220** | that is going to | 7(2.1%) |
| **221** | that the probability of | 7(2.1%) |
| **222** | that we need to | 7(2.1%) |
| **223** | the centre of the | 7(2.1%) |
| **224** | the last few years | 7(2.1%) |
| **225** | the structure of the | 7(2.1%) |
| **226** | the total number of | 7(2.1%) |
| **227** | then we're going to | 7(2.1%) |
| **228** | thing i want to | 7(2.1%) |
| **229** | this is what we | 7(2.1%) |
| **230** | this sort of thing | 7(2.1%) |
| **231** | to do now is | 7(2.1%) |
| **232** | we'll come back to | 7(2.1%) |
| **233** | when i was a | 7(2.1%) |
| **234** | when you look at | 7(2.1%) |
| **235** | who are going to | 7(2.1%) |
| **236** | you're going to have | 7(2.1%) |
| **237** | a lot of them | 6(1.8%) |
| **238** | a lot of these | 6(1.8%) |
| **239** | a problem with the | 6(1.8%) |
| **240** | a whole range of | 6(1.8%) |
| **241** | always going to be | 6(1.8%) |
| **242** | an awful lot of | 6(1.8%) |
| **243** | an example of a | 6(1.8%) |
| **244** | and i think that | 6(1.8%) |
| **245** | and i think that's | 6(1.8%) |

| | | |
|---|---|---|
| **246** | and look at the | 6(1.8%) |
| **247** | and they're going to | 6(1.8%) |
| **248** | and this is one | 6(1.8%) |
| **249** | and you need to | 6(1.8%) |
| **250** | are a number of | 6(1.8%) |
| **251** | are more likely to | 6(1.8%) |
| **252** | as a result of | 6(1.8%) |
| **253** | be able to do | 6(1.8%) |
| **254** | give you an example | 6(1.8%) |
| **255** | gives you an idea | 6(1.8%) |
| **256** | going to be in | 6(1.8%) |
| **257** | going to be talking | 6(1.8%) |
| **258** | going to give you | 6(1.8%) |
| **259** | going to have to | 6(1.8%) |
| **260** | have to be able | 6(1.8%) |
| **261** | i don't know i | 6(1.8%) |
| **262** | i don't know whether | 6(1.8%) |
| **263** | i think it's the | 6(1.8%) |
| **264** | i think we have | 6(1.8%) |
| **265** | i thought it was | 6(1.8%) |
| **266** | i want to say | 6(1.8%) |
| **267** | i'm going to be | 6(1.8%) |
| **268** | i'm going to give | 6(1.8%) |
| **269** | i'm going to say | 6(1.8%) |
| **270** | i'm going to show | 6(1.8%) |
| **271** | if we're looking at | 6(1.8%) |
| **272** | if you haven't got | 6(1.8%) |
| **273** | if you were to | 6(1.8%) |
| **274** | in a way that | 6(1.8%) |
| **275** | in front of you | 6(1.8%) |
| **276** | in terms of a | 6(1.8%) |
| **277** | in the sense that | 6(1.8%) |
| **278** | is not going to | 6(1.8%) |
| **279** | is not the only | 6(1.8%) |
| **280** | it may be that | 6(1.8%) |
| **281** | just bear in mind | 6(1.8%) |
| **282** | just to give you | 6(1.8%) |
| **283** | not going to get | 6(1.8%) |
| **284** | of course there are | 6(1.8%) |
| **285** | okay so what i'm | 6(1.8%) |
| **286** | over the last few | 6(1.8%) |
| **287** | per cent of all | 6(1.8%) |
| **288** | right at the beginning | 6(1.8%) |
| **289** | so how do we | 6(1.8%) |
| **290** | so in this case | 6(1.8%) |
| **291** | so one of the | 6(1.8%) |
| **292** | so we have to | 6(1.8%) |
| **293** | talking to you about | 6(1.8%) |
| **294** | that a lot of | 6(1.8%) |
| **295** | that i'm going to | 6(1.8%) |

| | | |
|---|---|---|
| **296** | that if you have | 6(1.8%) |
| **297** | that if you've got | 6(1.8%) |
| **298** | that this is a | 6(1.8%) |
| **299** | that we're going to | 6(1.8%) |
| **300** | that you need to | 6(1.8%) |
| **301** | the context of the | 6(1.8%) |
| **302** | the fact that the | 6(1.8%) |
| **303** | the idea is that | 6(1.8%) |
| **304** | the same sort of | 6(1.8%) |
| **305** | the sort of thing | 6(1.8%) |
| **306** | the top of the | 6(1.8%) |
| **307** | there are a number | 6(1.8%) |
| **308** | there are plenty of | 6(1.8%) |
| **309** | there may be a | 6(1.8%) |
| **310** | there's a number of | 6(1.8%) |
| **311** | this is just a | 6(1.8%) |
| **312** | to give you some | 6(1.8%) |
| **313** | to go to the | 6(1.8%) |
| **314** | to say something about | 6(1.8%) |
| **315** | want to get rid | 6(1.8%) |
| **316** | we can say that | 6(1.8%) |
| **317** | we talk about the | 6(1.8%) |
| **318** | we're going to do | 6(1.8%) |
| **319** | we're going to have | 6(1.8%) |
| **320** | we're going to start | 6(1.8%) |
| **321** | what happens to the | 6(1.8%) |
| **322** | what i'm talking about | 6(1.8%) |
| **323** | you an idea of | 6(1.8%) |
| **324** | you do need to | 6(1.8%) |
| **325** | you don't want to | 6(1.8%) |
| **326** | you go back to | 6(1.8%) |
| **327** | you have to do | 6(1.8%) |

**APPENDIX E: Shared bundles between Arts and Humanities and Life and Medical Sciences corpora:**

| No | Lexical bundles | Arts & Humanities | Life & Medical Sciences |
|---|---|---|---|
| **1** | the way in which | 77(24.9%) | 10(3%) |
| **2** | the end of the | 60(19.4%) | 36(11%) |
| **3** | at the end of | 49(15.8%) | 39(11.9%) |
| **4** | if you look at | 37(11.9%) | 24(7.3%) |
| **5** | on the other hand | 31(10%) | 13(3.9%) |
| **6** | what i want to | 31(10%) | 15(4.5%) |
| **7** | at the same time | 30(9.7%) | 17(5.1%) |
| **8** | in other words the | 30(9.7%) | 10(3%) |
| **9** | is one of the | 30(9.7%) | 12(3.6%) |
| **10** | and so on and | 28(9%) | 10(3%) |
| **11** | one of the things | 28(9%) | 26(7.9%) |
| **12** | in terms of the | 26(8.4%) | 15(4.5%) |
| **13** | and one of the | 25(8%) | 13(3.9%) |
| **14** | the rest of it | 23(7.4%) | 12(3.6%) |
| **15** | to be able to | 22(7.1%) | 47(14.3%) |
| **16** | at the beginning of | 21(6.7%) | 8(2.4%) |
| **17** | the rest of the | 21(6.7%) | 14(4.2%) |
| **18** | and you can see | 19(6.1%) | 35(10.7%) |
| **19** | so in other words | 19(6.1%) | 8(2.4%) |
| **20** | to do with the | 19(6.1%) | 11(3.3%) |
| **21** | a lot of the | 18(5.8%) | 15(4.5%) |
| **22** | and this is the | 18(5.8%) | 25(7.6%) |
| **23** | i don't know if | 18(5.8%) | 19(5.8%) |
| **24** | in the context of | 18(5.8%) | 10(3%) |
| **25** | you look at the | 18(5.8%) | 15(4.5%) |
| **26** | i want to do | 17(5.5%) | 8(2.4%) |
| **27** | and i'm going to | 16(5.1%) | 22(6.7%) |
| **28** | what i'm going to | 16(5.1%) | 19(5.8%) |
| **29** | if you want to | 15(4.8%) | 45(13.7%) |
| **30** | the nature of the | 15(4.8%) | 11(3.3%) |
| **31** | and this is a | 14(4.5%) | 19(5.8%) |
| **32** | i'm not going to | 14(4.5%) | 40(12.2%) |
| **33** | is going to be | 14(4.5%) | 33(10%) |
| **34** | those of you who | 14(4.5%) | 15(4.5%) |
| **35** | you can see that | 14(4.5%) | 41(12.5%) |
| **36** | are going to be | 13(4.2%) | 28(8.5%) |
| **37** | a little bit more | 12(3.8%) | 10(3%) |
| **38** | as you can see | 12(3.8%) | 13(3.9%) |
| **39** | to do is to | 12(3.8%) | 14(4.2%) |
| **40** | to look at the | 12(3.8%) | 13(3.9%) |
| **41** | i just want to | 11(3.5%) | 17(5.1%) |
| **42** | in the case of | 11(3.5%) | 36(11%) |
| **43** | or something like that | 11(3.5%) | 13(3.9%) |
| **44** | a little bit about | 10(3.2%) | 21(6.4%) |
| **45** | a little bit of | 10(3.2%) | 28(8.5%) |
| **46** | going to be the | 10(3.2%) | 11(3.3%) |

| | | | |
|---|---|---|---|
| **47** | going to look at | 10(3.2%) | 25(7.6%) |
| **48** | going to show you | 10(3.2%) | 7(2.1%) |
| **49** | going to talk about | 10(3.2%) | 48(14.6%) |
| **50** | i'll come back to | 10(3.2%) | 7(2.1%) |
| **51** | i'm going to show | 10(3.2%) | 6(1.8%) |
| **52** | in a way that | 10(3.2%) | 6(1.8%) |
| **53** | in the same way | 10(3.2%) | 8(2.4%) |
| **54** | it has to be | 10(3.2%) | 7(2.1%) |
| **55** | of the things that | 10(3.2%) | 23(7%) |
| **56** | and of course the | 9(2.9%) | 10(3%) |
| **57** | going to be a | 9(2.9%) | 14(4.2%) |
| **58** | i don't want to | 9(2.9%) | 8(2.4%) |
| **59** | i'm going to be | 9(2.9%) | 6(1.8%) |
| **60** | i'm going to do | 9(2.9%) | 11(3.3%) |
| **61** | in the middle of | 9(2.9%) | 11(3.3%) |
| **62** | so i'm going to | 9(2.9%) | 16(4.8%) |
| **63** | so that you can | 9(2.9%) | 8(2.4%) |
| **64** | to give you a | 9(2.9%) | 9(2.7%) |
| **65** | what we're going to | 9(2.9%) | 9(2.7%) |
| **66** | as a result of | 8(2.5%) | 6(1.8%) |
| **67** | by the end of | 8(2.5%) | 8(2.4%) |
| **68** | have a look at | 8(2.5%) | 7(2.1%) |
| **69** | i want you to | 8(2.5%) | 15(4.5%) |
| **70** | it's going to be | 8(2.5%) | 20(6.1%) |
| **71** | that i want to | 8(2.5%) | 7(2.1%) |
| **72** | the fact that the | 8(2.5%) | 6(1.8%) |
| **73** | to bear in mind | 8(2.5%) | 11(3.3%) |
| **74** | we're going to be | 8(2.5%) | 12(3.6%) |
| **75** | when we look at | 8(2.5%) | 11(3.3%) |
| **76** | and that is the | 7(2.2%) | 7(2.1%) |
| **77** | as well as the | 7(2.2%) | 7(2.1%) |
| **78** | come back to that | 7(2.2%) | 11(3.3%) |
| **79** | if we look at | 7(2.2%) | 26(7.9%) |
| **80** | in the sense that | 7(2.2%) | 6(1.8%) |
| **81** | it may be that | 7(2.2%) | 6(1.8%) |
| **82** | just to remind you | 7(2.2%) | 10(3%) |
| **83** | not going to be | 7(2.2%) | 11(3.3%) |
| **84** | of you who are | 7(2.2%) | 10(3%) |
| **85** | so this is the | 7(2.2%) | 14(4.2%) |
| **86** | so you can see | 7(2.2%) | 13(3.9%) |
| **87** | want to do is | 7(2.2%) | 8(2.4%) |
| **88** | you can see the | 7(2.2%) | 14(4.2%) |
| **89** | you're going to be | 7(2.2%) | 15(4.5%) |
| **90** | a bit of a | 6(1.9%) | 9(2.7%) |
| **91** | and this is what | 6(1.9%) | 13(3.9%) |
| **92** | but the point is | 6(1.9%) | 8(2.4%) |
| **93** | give you an example | 6(1.9%) | 6(1.8%) |
| **94** | i want to talk | 6(1.9%) | 7(2.1%) |
| **95** | i was going to | 6(1.9%) | 7(2.1%) |
| **96** | if you're going to | 6(1.9%) | 7(2.1%) |

| | | | |
|---|---|---|---|
| **97** | so there is a | 6(1.9%) | 7(2.1%) |
| **98** | so this is a | 6(1.9%) | 16(4.8%) |
| **99** | that a lot of | 6(1.9%) | 6(1.8%) |
| **100** | the top of the | 6(1.9%) | 6(1.8%) |
| **101** | there's a lot of | 6(1.9%) | 15(4.5%) |
| **102** | there's going to be | 6(1.9%) | 8(2.4%) |
| **103** | to do now is | 6(1.9%) | 7(2.1%) |
| **104** | we look at the | 6(1.9%) | 12(3.6%) |
| **105** | we're going to look | 6(1.9%) | 18(5.5%) |

**APPENDIX F: Lexical bundles specific to Arts and Humanities sub-corpus:**

| No | Lexical bundles | Frequency |
|---|---|---|
| **1** | on the one hand | 29(9.3%) |
| **2** | the idea of the | 24(7.7%) |
| **3** | all the rest of | 23(7.4%) |
| **4** | as a kind of | 23(7.4%) |
| **5** | and all the rest | 21(6.7%) |
| **6** | if you like the | 17(5.5%) |
| **7** | in the way that | 17(5.5%) |
| **8** | so on and so | 17(5.5%) |
| **9** | as a sort of | 16(5.1%) |
| **10** | is a kind of | 16(5.1%) |
| **11** | you know this is | 16(5.1%) |
| **12** | if you like of | 15(4.8%) |
| **13** | on and so forth | 15(4.8%) |
| **14** | one of the most | 15(4.8%) |
| **15** | at the heart of | 14(4.5%) |
| **16** | one of the reasons | 14(4.5%) |
| **17** | the beginning of the | 14(4.5%) |
| **18** | the idea of a | 14(4.5%) |
| **19** | way in which the | 14(4.5%) |
| **20** | the way that the | 13(4.2%) |
| **21** | a whole series of | 12(3.8%) |
| **22** | at the centre of | 12(3.8%) |
| **23** | at the start of | 12(3.8%) |
| **24** | is the kind of | 12(3.8%) |
| **25** | sort of you know | 12(3.8%) |
| **26** | the meaning of the | 12(3.8%) |
| **27** | the ways in which | 12(3.8%) |
| **28** | there's a kind of | 12(3.8%) |
| **29** | this is one of | 12(3.8%) |
| **30** | which i think is | 12(3.8%) |
| **31** | in so far as | 11(3.5%) |
| **32** | one of the great | 11(3.5%) |
| **33** | point of view of | 11(3.5%) |
| **34** | some of you may | 11(3.5%) |
| **35** | that there is a | 11(3.5%) |
| **36** | the interests of the | 11(3.5%) |
| **37** | the middle of the | 11(3.5%) |
| **38** | the point of view | 11(3.5%) |
| **39** | the whole of the | 11(3.5%) |
| **40** | was a kind of | 11(3.5%) |
| **41** | an idea of a | 10(3.2%) |
| **42** | at the level of | 10(3.2%) |
| **43** | for the first time | 10(3.2%) |
| **44** | i'm just going to | 10(3.2%) |
| **45** | if you like in | 10(3.2%) |
| **46** | in front of you | 10(3.2%) |
| **47** | in terms of what | 10(3.2%) |
| **48** | in the first place | 10(3.2%) |

| | | |
|---|---|---|
| **49** | is a sort of | 10(3.2%) |
| **50** | of the sort of | 10(3.2%) |
| **51** | of the things that | 10(3.2%) |
| **52** | the course of the | 10(3.2%) |
| **53** | the start of the | 10(3.2%) |
| **54** | the turn of the | 10(3.2%) |
| **55** | to the idea of | 10(3.2%) |
| **56** | was one of the | 10(3.2%) |
| **57** | way in which we | 10(3.2%) |
| **58** | a kind of a | 9(2.9%) |
| **59** | about the nature of | 9(2.9%) |
| **60** | at the top of | 9(2.9%) |
| **61** | draw your attention to | 9(2.9%) |
| **62** | from the point of | 9(2.9%) |
| **63** | i think it's a | 9(2.9%) |
| **64** | if you like to | 9(2.9%) |
| **65** | in a number of | 9(2.9%) |
| **66** | in a sort of | 9(2.9%) |
| **67** | in terms of a | 9(2.9%) |
| **68** | is the way in | 9(2.9%) |
| **69** | it seems to me | 9(2.9%) |
| **70** | it's a sort of | 9(2.9%) |
| **71** | of the kind of | 9(2.9%) |
| **72** | the idea that the | 9(2.9%) |
| **73** | what it is to | 9(2.9%) |
| **74** | what it means to | 9(2.9%) |
| **75** | about the way in | 8(2.5%) |
| **76** | and as i say | 8(2.5%) |
| **77** | and the idea of | 8(2.5%) |
| **78** | and the way in | 8(2.5%) |
| **79** | but i think it's | 8(2.5%) |
| **80** | for those of you | 8(2.5%) |
| **81** | i mean this is | 8(2.5%) |
| **82** | if you think of | 8(2.5%) |
| **83** | in a kind of | 8(2.5%) |
| **84** | in a way which | 8(2.5%) |
| **85** | in relation to the | 8(2.5%) |
| **86** | in this sort of | 8(2.5%) |
| **87** | is part of the | 8(2.5%) |
| **88** | of the reasons why | 8(2.5%) |
| **89** | that is to say | 8(2.5%) |
| **90** | that kind of thing | 8(2.5%) |
| **91** | the back of the | 8(2.5%) |
| **92** | the one of the | 8(2.5%) |
| **93** | the second half of | 8(2.5%) |
| **94** | to a certain extent | 8(2.5%) |
| **95** | to come back to | 8(2.5%) |
| **96** | to think about the | 8(2.5%) |
| **97** | towards the end of | 8(2.5%) |
| **98** | what's going on in | 8(2.5%) |

| | | |
|---|---|---|
| **99** | you know in the | 8(2.5%) |
| **100** | you might want to | 8(2.5%) |
| **101** | a great deal of | 7(2.2%) |
| **102** | a way in which | 7(2.2%) |
| **103** | and i want to | 7(2.2%) |
| **104** | and on the other | 7(2.2%) |
| **105** | and so on but | 7(2.2%) |
| **106** | and the fact that | 7(2.2%) |
| **107** | come back to this | 7(2.2%) |
| **108** | going on in the | 7(2.2%) |
| **109** | have to look at | 7(2.2%) |
| **110** | i think this is | 7(2.2%) |
| **111** | if you like and | 7(2.2%) |
| **112** | in a sense the | 7(2.2%) |
| **113** | in such a way | 7(2.2%) |
| **114** | in the course of | 7(2.2%) |
| **115** | in the sort of | 7(2.2%) |
| **116** | is in fact a | 7(2.2%) |
| **117** | is the idea that | 7(2.2%) |
| **118** | it in terms of | 7(2.2%) |
| **119** | it's not just that | 7(2.2%) |
| **120** | now i want to | 7(2.2%) |
| **121** | of the things which | 7(2.2%) |
| **122** | of the ways in | 7(2.2%) |
| **123** | second half of the | 7(2.2%) |
| **124** | something to do with | 7(2.2%) |
| **125** | that in a minute | 7(2.2%) |
| **126** | that's one of the | 7(2.2%) |
| **127** | the fact that he | 7(2.2%) |
| **128** | the fact that it | 7(2.2%) |
| **129** | the first part of | 7(2.2%) |
| **130** | the heart of the | 7(2.2%) |
| **131** | the people who are | 7(2.2%) |
| **132** | the same kind of | 7(2.2%) |
| **133** | this is the kind | 7(2.2%) |
| **134** | to say that the | 7(2.2%) |
| **135** | to the fact that | 7(2.2%) |
| **136** | was going to be | 7(2.2%) |
| **137** | you have to remember | 7(2.2%) |
| **138** | you know if you | 7(2.2%) |
| **139** | you look at it | 7(2.2%) |
| **140** | a certain amount of | 6(1.9%) |
| **141** | a sort of a | 6(1.9%) |
| **142** | all sorts of things | 6(1.9%) |
| **143** | and he had a | 6(1.9%) |
| **144** | and i think the | 6(1.9%) |
| **145** | and i think this | 6(1.9%) |
| **146** | and that is a | 6(1.9%) |
| **147** | and that is that | 6(1.9%) |
| **148** | and think about the | 6(1.9%) |

| | | |
|---|---|---|
| **149** | are we going to | 6(1.9%) |
| **150** | as it were a | 6(1.9%) |
| **151** | as well and the | 6(1.9%) |
| **152** | at some of the | 6(1.9%) |
| **153** | at the bottom of | 6(1.9%) |
| **154** | at the time and | 6(1.9%) |
| **155** | but i think the | 6(1.9%) |
| **156** | but on the other | 6(1.9%) |
| **157** | can be seen as | 6(1.9%) |
| **158** | first of all the | 6(1.9%) |
| **159** | for the rest of | 6(1.9%) |
| **160** | i just wanted to | 6(1.9%) |
| **161** | i said at the | 6(1.9%) |
| **162** | i think you know | 6(1.9%) |
| **163** | i was talking about | 6(1.9%) |
| **164** | if you like which | 6(1.9%) |
| **165** | in one of his | 6(1.9%) |
| **166** | in which we think | 6(1.9%) |
| **167** | is to do with | 6(1.9%) |
| **168** | kind of you know | 6(1.9%) |
| **169** | know this is a | 6(1.9%) |
| **170** | look at some of | 6(1.9%) |
| **171** | now what i want | 6(1.9%) |
| **172** | of the fact that | 6(1.9%) |
| **173** | of this sort of | 6(1.9%) |
| **174** | of what's going on | 6(1.9%) |
| **175** | of you know the | 6(1.9%) |
| **176** | on the part of | 6(1.9%) |
| **177** | on the top of | 6(1.9%) |
| **178** | one way or another | 6(1.9%) |
| **179** | seems to me that | 6(1.9%) |
| **180** | so far as it | 6(1.9%) |
| **181** | that there was a | 6(1.9%) |
| **182** | the bottom of the | 6(1.9%) |
| **183** | the centre of the | 6(1.9%) |
| **184** | the kind of the | 6(1.9%) |
| **185** | the one hand and | 6(1.9%) |
| **186** | the role of the | 6(1.9%) |
| **187** | the second part of | 6(1.9%) |
| **188** | the way that it | 6(1.9%) |
| **189** | the you know the | 6(1.9%) |
| **190** | there will be a | 6(1.9%) |
| **191** | there would be a | 6(1.9%) |
| **192** | there's a sort of | 6(1.9%) |
| **193** | this idea of the | 6(1.9%) |
| **194** | this is something that | 6(1.9%) |
| **195** | this is what the | 6(1.9%) |
| **196** | to be looking at | 6(1.9%) |
| **197** | to look at this | 6(1.9%) |
| **198** | to look at what | 6(1.9%) |

| | | |
|---|---|---|
| **199** | to talk about the | 6(1.9%) |
| **200** | trying to do is | 6(1.9%) |
| **201** | want to think about | 6(1.9%) |
| **202** | we'll be looking at | 6(1.9%) |
| **203** | what does it mean | 6(1.9%) |
| **204** | you like of the | 6(1.9%) |

**APPENDIX G: Lexical bundles specific to Life and Medical Sciences sub-corpus:**

| No | Lexical bundles | Frequency |
|---|---|---|
| **1** | per cent of the | 45(13.7%) |
| **2** | i'm going to talk | 30(9.1%) |
| **3** | and things like that | 23(7%) |
| **4** | thank you very much | 20(6.1%) |
| **5** | per cent of people | 19(5.8%) |
| **6** | if you have a | 16(4.8%) |
| **7** | quite a lot of | 16(4.8%) |
| **8** | so you have to | 15(4.5%) |
| **9** | what do you think | 15(4.5%) |
| **10** | you can see here | 15(4.5%) |
| **11** | you need to know | 15(4.5%) |
| **12** | can see that the | 14(4.2%) |
| **13** | if you think about | 14(4.2%) |
| **14** | not going to go | 14(4.2%) |
| **15** | but you can see | 13(3.9%) |
| **16** | okay so this is | 13(3.9%) |
| **17** | to get rid of | 13(3.9%) |
| **18** | you don't need to | 13(3.9%) |
| **19** | you think about it | 13(3.9%) |
| **20** | you're not going to | 13(3.9%) |
| **21** | a lot of people | 12(3.6%) |
| **22** | and if you look | 12(3.6%) |
| **23** | and we're going to | 12(3.6%) |
| **24** | go back to the | 12(3.6%) |
| **25** | going to go into | 12(3.6%) |
| **26** | this part of the | 12(3.6%) |
| **27** | to talk about the | 12(3.6%) |
| **28** | why do you think | 12(3.6%) |
| **29** | are you going to | 11(3.3%) |
| **30** | for a long time | 11(3.3%) |
| **31** | over a period of | 11(3.3%) |
| **32** | so if we look | 11(3.3%) |
| **33** | we're going to talk | 11(3.3%) |
| **34** | you can see is | 11(3.3%) |
| **35** | you have to have | 11(3.3%) |
| **36** | a bit more about | 10(3%) |
| **37** | and these are the | 10(3%) |
| **38** | and you have to | 10(3%) |
| **39** | are there any questions | 10(3%) |
| **40** | for those of you | 10(3%) |
| **41** | if you don't know | 10(3%) |
| **42** | talk a little bit | 10(3%) |
| **43** | what i mean by | 10(3%) |
| **44** | what's going to happen | 10(3%) |
| **45** | and this is because | 9(2.7%) |
| **46** | are a lot of | 9(2.7%) |
| **47** | at the back there | 9(2.7%) |
| **48** | bear that in mind | 9(2.7%) |

| | | |
|---|---|---|
| **49** | if you've got a | 9(2.7%) |
| **50** | in the absence of | 9(2.7%) |
| **51** | it's one of the | 9(2.7%) |
| **52** | it's very difficult to | 9(2.7%) |
| **53** | nothing to do with | 9(2.7%) |
| **54** | one of the most | 9(2.7%) |
| **55** | one of the problems | 9(2.7%) |
| **56** | so if you have | 9(2.7%) |
| **57** | so these are the | 9(2.7%) |
| **58** | so we need to | 9(2.7%) |
| **59** | that you're going to | 9(2.7%) |
| **60** | there are a lot | 9(2.7%) |
| **61** | there are lots of | 9(2.7%) |
| **62** | they're going to be | 9(2.7%) |
| **63** | this is one of | 9(2.7%) |
| **64** | want to talk about | 9(2.7%) |
| **65** | we're not going to | 9(2.7%) |
| **66** | you've got to have | 9(2.7%) |
| **67** | and a lot of | 8(2.4%) |
| **68** | and i don't know | 8(2.4%) |
| **69** | as we go along | 8(2.4%) |
| **70** | at the moment and | 8(2.4%) |
| **71** | do you think it's | 8(2.4%) |
| **72** | i don't know how | 8(2.4%) |
| **73** | i don't know what | 8(2.4%) |
| **74** | i don't think it's | 8(2.4%) |
| **75** | i said to you | 8(2.4%) |
| **76** | in any great detail | 8(2.4%) |
| **77** | is that if you | 8(2.4%) |
| **78** | not going to talk | 8(2.4%) |
| **79** | one of the reasons | 8(2.4%) |
| **80** | per cent of your | 8(2.4%) |
| **81** | so if we have | 8(2.4%) |
| **82** | so if you want | 8(2.4%) |
| **83** | so what i'm going | 8(2.4%) |
| **84** | that one of the | 8(2.4%) |
| **85** | the first thing that | 8(2.4%) |
| **86** | the most important thing | 8(2.4%) |
| **87** | the size of the | 8(2.4%) |
| **88** | those of you that | 8(2.4%) |
| **89** | to make sure that | 8(2.4%) |
| **90** | to talk to you | 8(2.4%) |
| **91** | we have to be | 8(2.4%) |
| **92** | when it comes to | 8(2.4%) |
| **93** | you don't have to | 8(2.4%) |
| **94** | you end up with | 8(2.4%) |
| **95** | you're going to get | 8(2.4%) |
| **96** | a bit more detail | 7(2.1%) |
| **97** | a long period of | 7(2.1%) |
| **98** | a small amount of | 7(2.1%) |

| | | |
|---|---|---|
| **99** | and i think the | 7(2.1%) |
| **100** | and if you do | 7(2.1%) |
| **101** | and if you don't | 7(2.1%) |
| **102** | and if you have | 7(2.1%) |
| **103** | and if you think | 7(2.1%) |
| **104** | and there is a | 7(2.1%) |
| **105** | be a little bit | 7(2.1%) |
| **106** | by the time you | 7(2.1%) |
| **107** | come and see me | 7(2.1%) |
| **108** | do you think the | 7(2.1%) |
| **109** | does that make sense | 7(2.1%) |
| **110** | get rid of the | 7(2.1%) |
| **111** | if we're going to | 7(2.1%) |
| **112** | if you do a | 7(2.1%) |
| **113** | if you forget everything | 7(2.1%) |
| **114** | in a bit more | 7(2.1%) |
| **115** | in other words you | 7(2.1%) |
| **116** | is a little bit | 7(2.1%) |
| **117** | is the fact that | 7(2.1%) |
| **118** | is the sort of | 7(2.1%) |
| **119** | it used to be | 7(2.1%) |
| **120** | little bit about the | 7(2.1%) |
| **121** | most of the time | 7(2.1%) |
| **122** | now i'm going to | 7(2.1%) |
| **123** | so here we have | 7(2.1%) |
| **124** | so you need to | 7(2.1%) |
| **125** | so you've got a | 7(2.1%) |
| **126** | that is going to | 7(2.1%) |
| **127** | that the probability of | 7(2.1%) |
| **128** | that we need to | 7(2.1%) |
| **129** | the centre of the | 7(2.1%) |
| **130** | the last few years | 7(2.1%) |
| **131** | the structure of the | 7(2.1%) |
| **132** | the total number of | 7(2.1%) |
| **133** | then we're going to | 7(2.1%) |
| **134** | thing i want to | 7(2.1%) |
| **135** | this is what we | 7(2.1%) |
| **136** | this sort of thing | 7(2.1%) |
| **137** | we'll come back to | 7(2.1%) |
| **138** | when i was a | 7(2.1%) |
| **139** | when you look at | 7(2.1%) |
| **140** | who are going to | 7(2.1%) |
| **141** | you're going to have | 7(2.1%) |
| **142** | a lot of them | 6(1.8%) |
| **143** | a lot of these | 6(1.8%) |
| **144** | a problem with the | 6(1.8%) |
| **145** | a whole range of | 6(1.8%) |
| **146** | always going to be | 6(1.8%) |
| **147** | an awful lot of | 6(1.8%) |
| **148** | an example of a | 6(1.8%) |

| | | |
|---|---|---|
| **149** | and i think that | 6(1.8%) |
| **150** | and i think that's | 6(1.8%) |
| **151** | and look at the | 6(1.8%) |
| **152** | and they're going to | 6(1.8%) |
| **153** | and this is one | 6(1.8%) |
| **154** | and you need to | 6(1.8%) |
| **155** | are a number of | 6(1.8%) |
| **156** | are more likely to | 6(1.8%) |
| **157** | be able to do | 6(1.8%) |
| **158** | gives you an idea | 6(1.8%) |
| **159** | going to be in | 6(1.8%) |
| **160** | going to be talking | 6(1.8%) |
| **161** | going to give you | 6(1.8%) |
| **162** | going to have to | 6(1.8%) |
| **163** | have to be able | 6(1.8%) |
| **164** | i don't know i | 6(1.8%) |
| **165** | i don't know whether | 6(1.8%) |
| **166** | i think it's the | 6(1.8%) |
| **167** | i think we have | 6(1.8%) |
| **168** | i thought it was | 6(1.8%) |
| **169** | i want to say | 6(1.8%) |
| **170** | i'm going to give | 6(1.8%) |
| **171** | i'm going to say | 6(1.8%) |
| **172** | if we're looking at | 6(1.8%) |
| **173** | if you haven't got | 6(1.8%) |
| **174** | if you were to | 6(1.8%) |
| **175** | in front of you | 6(1.8%) |
| **176** | in terms of a | 6(1.8%) |
| **177** | is not going to | 6(1.8%) |
| **178** | is not the only | 6(1.8%) |
| **179** | just bear in mind | 6(1.8%) |
| **180** | just to give you | 6(1.8%) |
| **181** | not going to get | 6(1.8%) |
| **182** | of course there are | 6(1.8%) |
| **183** | okay so what i'm | 6(1.8%) |
| **184** | over the last few | 6(1.8%) |
| **185** | per cent of all | 6(1.8%) |
| **186** | right at the beginning | 6(1.8%) |
| **187** | so how do we | 6(1.8%) |
| **188** | so in this case | 6(1.8%) |
| **189** | so one of the | 6(1.8%) |
| **190** | so we have to | 6(1.8%) |
| **191** | talking to you about | 6(1.8%) |
| **192** | that i'm going to | 6(1.8%) |
| **193** | that if you have | 6(1.8%) |
| **194** | that if you've got | 6(1.8%) |
| **195** | that this is a | 6(1.8%) |
| **196** | that we're going to | 6(1.8%) |
| **197** | that you need to | 6(1.8%) |
| **198** | the context of the | 6(1.8%) |

| | | |
|---|---|---|
| **199** | the idea is that | 6(1.8%) |
| **200** | the same sort of | 6(1.8%) |
| **201** | the sort of thing | 6(1.8%) |
| **202** | there are a number | 6(1.8%) |
| **203** | there are plenty of | 6(1.8%) |
| **204** | there may be a | 6(1.8%) |
| **205** | there's a number of | 6(1.8%) |
| **206** | this is just a | 6(1.8%) |
| **207** | to give you some | 6(1.8%) |
| **208** | to go to the | 6(1.8%) |
| **209** | to say something about | 6(1.8%) |
| **210** | want to get rid | 6(1.8%) |
| **211** | we can say that | 6(1.8%) |
| **212** | we talk about the | 6(1.8%) |
| **213** | we're going to start | 6(1.8%) |
| **214** | what happens to the | 6(1.8%) |
| **215** | what i'm talking about | 6(1.8%) |
| **216** | you an idea of | 6(1.8%) |
| **217** | you do need to | 6(1.8%) |
| **218** | you go back to | 6(1.8%) |

**APPENDIX H: Functional distribution of lexical bundles in Arts and Humanities lectures**

| Functional categories | Lexical bundles |
|---|---|
| **1. Stance expressions** | |
| A. Epistemic stance | |
| *Personal* | i don't know if, which i think is, i think it's a, it seems to me, i think this is, and i think the, and i think this, but i think the, i think you know, seems to me that |
| *Impersonal* | the fact that the, and the fact that, is in fact a, the fact that he, the fact that it, to the fact that, of the fact that |
| B. Attitudinal/ modality stance | |
| *B1. Desire* | |
| *Personal* | if you want to, i just want to, i don't want to, now i want to, i just wanted to |
| *Impersonal* | ------ |
| *B2. Obligation/ directive* | |
| *Personal* | i want you to, you might want to, you have to remember you look at the |
| *Impersonal* | it has to be |
| *B3. Intention/ prediction* | |
| *Personal* | and i'm going to, i'm not going to, is going to be, are going to be, going to show you, i'm going to show, i'm just going to, i'm going to be, i'm going to do, so i'm going to, to give you a, we're going to be, and i want to, was going to be, you're going to be, are we going to, i was going to, if you're going to |
| *Impersonal* | going to be the, going to be a, it's going to be, going on in the, not going to be, want to do is, there's going to be |
| *B4. Ability* | |
| *Personal* | to be able to, so that you can |
| *Impersonal* | ------ |
| **2. Discourse organizers** | |
| A. Topic introduction/focus | if you look at, what i want to, the idea of the, and you can see, i want to do, if you like the, what i'm going to, if you like of, the idea of a, you can see that, to look at the, a little bit about, an idea of a, going to look at, going to talk about, i'll come back to, if you like in, to the idea of, draw your attention to, from the point of, if you like to, the idea that the, what it is to, |

| | |
|---|---|
| | what we're going to, and the idea of, have a look at, if you think of, that i want to, to bear in mind, to come back to, to think about the, what's going on in, when we look at, come back to that, come back to this, have to look at, if we look at, if you like and, is the idea that, it may be that, just to remind you, so you can see, you can see the, you look at it, and he had a, and think about the, as it were a, first of all the, give you an example, i said at the, i want to talk, i was talking about, if you like which, in which we think, kind of you know, look at some of, now what i want, of what's going on, the you know the, there will be a, there would be a, this idea of the, this is what the, to be looking at, to do now is, to look at this, to look at what, to talk about the, trying to do is, want to think about, we look at the, we'll be looking at, we're going to look, what does it mean, you like of the, to do is to |
| B. Topic elaboration/ clarification | on the other hand, in other words the, on the one hand, so in other words, to do with the, you know this is, on and so forth, as you can see, sort of you know, in so far as, point of view of, the point of view, in the same way, of course in the, and of course the, what it means to, and as i say, but i think it's, i mean this is, of the reasons why, that is to say, you know in the, and on the other, as well as the, in such a way, something to do with, to say that the, you know if you, as well and the, but on the other, can be seen as, is to do with, of you know the, one way or another, so far as it, the one hand and, but the point is |
| **3. Referential expressions** | |
| A. Identification/focus | is one of the, one of the things, and one of the, and this is the, in the way that, as a sort of, is a kind of, one of the most, and this is a, one of the reasons, those of you who, the way that the, is the kind of, there's a kind of, this is one of, one of the great, some of you may, that there is a, was a kind of, in a way that, is a sort of, of the sort of, of the things that, was one of the, a kind of a, in a sort of, is the way in, it's a sort of, of the kind of, about the way in, and the way in, for those of you, in a kind of, in a way which, in this sort of, is part of the, that kind of thing, the one of the, a way in which, and that is the, in the sort of, it's not just that, of the things which, of the ways in, so this is the, that's one of the, the first part of, the people who are, the same kind of, this is the kind, a bit of a, a sort of a, all sorts of things, and that is a, and that is that, and this is what, in one of his, know this is a, of this sort of, on the part of, so there is a, so this is a, that there was a, the kind of the |

| | |
|---|---|
| | the second part of, the way that it, there's a sort of, this is something that, as a kind of |
| B. Imprecision | and so on and, so on and so, or something like that, and so on but, and so forth and |
| C. Specification of attributes | |
| *C1. Quantity specification* | all the rest of, the rest of it, and all the rest, the rest of the, a lot of the, a little bit more, a whole series of, the whole of the, a little bit of, in a number of, to a certain extent, a great deal of, a certain amount of, at some of the, for the rest of, that a lot of, there's a lot of |
| *C2. Tangible framing attributes* | -------- |
| *C3. Intangible framing attributes* | in the context of, the course of the, in relation to the, in the course of, the way in which ,in terms of the, the nature of the, way in which the, the meaning of the, the ways in which, in the case of, the interests of the, at the level of, in terms of what, way in which we, about the nature of, in terms of a, as a result of, in a sense the, in the sense that, it in terms of, the heart of the, the role of the |
| D. Time/place/text reference | |
| *D1. Place reference* | at the centre of, at the start of, the middle of the, in front of you, the start of the, at the top of, at the bottom of, on the top of, the bottom of the, the top of the |
| *D2. Time reference* | at the same time, for the first time, in the first place, the turn of the, towards the end of, that in a minute, at the time and |
| *D3. Topic/ subject/ lecture reference* | i'll come back to, to come back to, come back to that, come back to this, we'll be looking at, i was talking about, i said at the |
| *D4. Multi-functional reference* | the end of the, at the end of, at the beginning of, at the heart of, the beginning of the, in the middle of, by the end of, the back of the, the second half of, second half of the, the centre of the |

| | |
|---|---|
| **4. Special conversational expressions** | |
| A. Politeness | ------ |
| B. Direct inquiry | ------ |
| C. Reporting | |

**Appendix I: Functional distribution of lexical bundles in Life and Medical Sciences**

| Functional categories | Lexical bundles |
|---|---|
| **1. Stance expressions** | |
| A. Epistemic stance | |
| *Personal* | i don't know if, and i don't know, i don't know how, i don't know what, i don't think it's, and i think the, and i think that, and i think that's, i don't know I, i don't know whether, i think it's the, i think we have, i thought it was |
| *Impersonal* | is the fact that, are more likely to, the fact that the |
| B. Attitudinal/modality stance | |
| *B1. Desire* | |
| *Personal* | i just want to, i don't want to, so if you want |
| *Impersonal* | -------- |
| *B2. Obligation/ directive* | |
| *Personal* | i want you to, so you have to, you look at the, you need to know, you don't need to, you have to have, and you have to, you can see here, so we need to, we have to be, so you need to, that we need to, you're going to have, and you need to, so we have to, that you need to, you do need to, you don't want to, you have to do, come and see me, if you don't know, you don't have to |
| *Impersonal* | it has to be |
| *B3. Intention/ prediction* | |
| *Personal* | i'm not going to, and i'm going to, so i'm going to, you're going to be, you're not going to, and we're going to, we're going to be, i'm going to do, are you going to, that you're going to, they're going to be, to give you a, we're not going to, what we're going to, you end up with, you're going to get, going to show you, i was going to, if we're going to, if you're going to, now i'm going to, thing i want to, and they're going to, going to give you, going to have to, i'm going to be, i'm going to give, i'm going to show, just to give you, that i'm going to, that we're going to, to give you some, we're going to have, we're going to start |
| *Impersonal* | is going to be, are going to be, it's going to be, going to be a, not going to go, going to go into, going to be the, not going to be, that is going to, always going to be, going to be in, is not going to, there's going to be, not going to get |
| *B4. Ability* | |
| *Personal* | so that you can, to be able to, be able to do, have to be able |
| *Impersonal* | ----- |

| | |
|---|---|
| **2. Discourse organizers**<br>A. Topic introduction/focus | going to talk about, if you want to, you can see that, and you can see, i'm going to talk, if we look at, going to look at, if you look at, a little bit about, what i'm going to, we're going to look, if you have a, what do you think, what i want to, can see that the, if you think about, to do is to, you can see the, as you can see, but you can see, so you can see, to look at the, you think about it, and if you look, go back to the, to talk about the, we look at the, why do you think, come back to that, so if we look, we're going to talk, when we look at, you can see is, a bit more about, talk a little bit, what's going to happen, bear that in mind, if you've got a, it's very difficult to, so if you have, want to talk about, you've got to have, as we go along, do you think it's, i want to do, in any great detail, not going to talk, so if we have, so what i'm going, to make sure that, to talk to you, want to do is, and if you do, and if you don't, and if you have, and if you think, do you think the, does that make sense, get rid of the, have a look at, i want to talk, i'll come back to, if you do a, if you forget everything, it used to be, little bit about the, so here we have, so you've got a, that i want to, then we're going to, this is what we, to do now is, we'll come back to, when i was a, when you look at, who are going to, and look at the, give you an example, gives you an idea, going to be talking, if we're looking at, if you haven't got, if you were to, it may be that, just bear in mind, okay so what i'm, right at the beginning, so how do we, talking to you about, that if you have, that if you've got, the idea is that, to go to the, want to get rid, we talk about the, we're going to do, what happens to the, what i'm talking about, you an idea of, you go back to, to get rid of, to bear in mind, just to remind you, when it comes to |
| B. Topic elaboration/clarification | on the other hand, to do with the, and of course the, in other words the, what i mean by, and this is because, nothing to do with, in the same way, is that if you, so in other words, as well as the, in other words you, a problem with the, i want to say, i'm going to say, of course there are, so in this case, to say something about, we can say that, , but the point is |
| **3. Referential expressions**<br>A. Identification/focus | one of the things, and this is the, of the things that, and this is a, so this is a, those of you who, so this is the, and one of the, and this is what, okay so this is, is one of the, and these are the, for those of you, of you who are, a bit of a, it's one of the, one of the most, one of the problems, so these are the, this is one of, one of the reasons, that one of the, the first thing that, the most important thing, those of you that, and that is the, and there is a, is the sort of, so there is a, this sort of thing, an example of a, and this is one, in a way that, is |

| | |
|---|---|
| | not the only, so one of the, that this is a, the same sort of, the sort of thing, there may be a, this is just a |
| B. Imprecision | and things like that, or something like that, and so on and |
| C. Specification of attributes | |
| *C1. Quantity specification* | per cent of the, a little bit of, per cent of people, quite a lot of, a lot of the, there's a lot of, the rest of the, a lot of people, the rest of it, a little bit more, are a lot of, there are a lot, there are lots of, and a lot of, per cent of your, a bit more detail, a small amount of, be a little bit, in a bit more, is a little bit, the total number of, a lot of them, a lot of these, a whole range of, an awful lot of, are a number of, per cent of all, that a lot of, there are a number, there are plenty of, there's a number of |
| *C2. Tangible framing attributes* | this part of the, the size of the, the structure of the |
| *C3. Intangible framing attributes* | in the case of, in terms of the, the nature of the, in the context of, the way in which, in the absence of, as a result of, in terms of a, in the sense that, the context of the, that the probability of |
| D. Time/place/text reference | |
| *D1. Place reference* | at the back there, in front of you, the top of the, the centre of the |
| *D2. Time reference* | at the same time, for a long time, over a period of, at the beginning of, at the moment and, a long period of, by the time you, most of the time, the last few years, over the last few |
| *D3. Topic/ subject/ lecture reference* | go back to the, come back to that, i'll come back to, we'll come back to |
| *D4. Multi-functional reference* | at the end of, the end of the, in the middle of, by the end of |
| **4. Special conversational functions** | |
| A. Politeness | thank you very much, |
| B. Simple inquiry | what do you think, does that make sense, are there any questions |
| C. Reporting | i said to you |