



UNIVERSITI PUTRA MALAYSIA

***CORPUS-BASED STUDY OF LEXICAL BUNDLES IN ACADEMIC
LECTURES ACROSS THREE DISCIPLINARY DIVISIONS***

HADI KASHIHA

FBMK 2015 93



**CORPUS-BASED STUDY OF LEXICAL BUNDLES IN ACADEMIC
LECTURES ACROSS THREE DISCIPLINARY DIVISIONS**

By

HADI KASHIHA

**Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia,
in Fulfillment of the Requirement for the Degree of Doctor of Philosophy**

June 2015

All material contained within the thesis, including without limitation text, logos, icons, photographs and all other artwork, is copyright material of Universiti Putra Malaysia unless otherwise stated. Use may be made of any material contained within the thesis for non-commercial purposes from the copyright holder. Commercial use of material may only be made with the express, prior, written permission of Universiti Putra Malaysia.

Copyright © Universiti Putra Malaysia



Specially Dedicated to:

MY late father



Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment
of the requirement for the degree of Doctor of Philosophy

CORPUS-BASED STUDY OF LEXICAL BUNDLES IN ACADEMIC LECTURES ACROSS THREE DISCIPLINARY DIVISIONS

By

HADI KASHIHA

June 2015

Chairperson: Prof. Chan Swee Heng, PhD
Faculty: Modern Languages and Communication

Simply defined as extended collocations, lexical bundles are combination of more than two words which co-occur frequently in a given register (Biber et al., 1999). They lead to coherence in text or speech 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 frequency, structure and discourse function of the most frequently occurring four-word lexical bundles in academic lectures across three broad disciplinary groupings, namely social sciences, physical sciences and life and medical sciences.

This comparative study was run on the nearly one million word corpus of 120 academic lectures (40 from each science). The lectures were transactional in nature and sourced from British Academic Spoken English (BASE) corpus. The most frequent four-word bundles were identified in each corpus using the computer program WordSmith Tools 5 (Scott, 2008). Then, the structural and functional taxonomies proposed by Biber et al. (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 three sciences in relation to the distributional patterns of the target bundles. In addition, the three 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 lecturers 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. The implication of this study direct itself to the

novice (especially those working in ESL/EFL settings) academic lecturers belonging to sciences under investigation. Findings of this study open more windows to how lexical bundles and their communicative functions are employed in academic disciplinary lectures. Students who study in these sciences could also benefit from findings of this research by being familiarized with the structural and functional characteristics of lexical bundles.

Key words: Academic lecture, disciplinary divisions, discourse function, formulaic language, lexical bundle



Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

ANALISIS BERASASKAN KORPUS TERHADAP RANGKAIAN LEKSIKAL DALAM SYARAHAN AKADEMIK MERENTAS TIGA DISIPLIN

Oleh

HADI KASHIHA

Jun 2015

Pengerusi: Prof. Chan Swee Heng, PhD
Fakulti: Bahasa Moden dan Komunikasi

Rangkaian leksikal merupakan kombinasi lebih daripada dua perkataan yang hadir bergandingan dalam daftar tertentu (Biber et. al., 1999). Penggunaan rangkaian leksikal dalam wacana teks atau ucapan penting dalam memastikan kefasihan produk linguistik, baik dalam bentuk tulisan mahupun lisan. Walaupun kajian terhadap rangkaian leksikal sudah banyak mendapat perhatian, masih belum dapat dipastikan tentang perbezaan penggunaannya dalam syarahan akademik yang berbeza disiplin. Bertitik tolak daripada persoalan tersebut, kajian ini akan mengenal pasti kekerapan, struktur dan fungsi rangkaian leksikal yang lazim digunakan dalam syarahan akademik merentas tiga kumpulan disiplin ilmu, iaitu disiplin sains sosial, sains fizik dan sains hayat dan perubatan.

Kajian ini dijalankan secara perbandingan ke atas hampir satu juta korpus kata daripada 120 syarahan akademik (40 bagi setiap disiplin ilmu). Syarahan akademik terpilih terdiri daripada syarahan kuliah bersumberkan korpus *British Academic Spoken English* (BASE). Empat rangkai kata paling kerap akan dikenal pasti dalam setiap korpus menggunakan program WordSmith Tools 5 (Scott, 2008). Kajian ini menggunakan taksonomi struktural dan fungsional yang dikemukakan oleh Biber et al., (2004) sebagai kerangka analisis Taksonomi Biber et al., (2014) diterapkan ketika pengelompokan rangkaian leksikal berdasarkan jenis dan fungsinya.

Dapatan kajian ini menemui beberapa perbezaan antara ketiga-tiga disiplin behubung dengan pola penyebaran rangkaian leksikal sasaran. Tiga daripada kumpulan syarahan turut menunjukkan perbezaan dari segi pemilihan jenis gramatikal yang digunakan. Dapatan kajian ini menunjukkan bahawa pemilihan rangkaian leksikal dan penggunaannya berdasarkan fungsi yang bersesuaian terhad dalam disiplin tertentu. Syarahan akademik didapati mempunyai kaedah pemilihan khusus yang sesuai dengan jurusan ilmu dan penerimaan pendengar. Berdasarkan dapatan ini, rangkaian leksikal didapati berpotensi untuk dijadikan indikator pembeza bagi syarahan akademik daripada disiplin berlainan. Dapatan kajian ini memberi sumbangan khusus kepada para pensyarah dan pelajar dalam disiplin yang terlibat. Hasil kajian ini terbatas pada skop yang diberi perhatian dan terbuka untuk kajian lanjutan oleh pengkaji akan datang.

Kata kunci: Syarahan akademik, disiplin, fungsi wacana, bahasa formulaik, rangkaian leksikal



ACKNOWLEDGEMENTS

There are many people I wish to thank, without whose help and support this thesis would not have been possible. First and foremost, my heartfelt thanks go to my supervisor **Prof. Dr. Chan Swee Heng** for sharing her pearls of wisdom with me and her useful comments, remarks and engagement during the course of this research. I also wish to express my sincere thanks to my committee members **Assoc. Dr. Ain Nadzimah Abdullah** and **Dr. Helen Tan**. They have given me invaluable help, guidance, and support throughout the course of the research and the writing of the thesis.

My deepest gratitude goes to my family and my friends. Without their love and support, I would not have overcome the difficulties encountered while conducting this research and writing this thesis.

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfillment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Chan Swee Heng, PhD

Professor

Faculty of Modern Languages and Communication

Universiti Putra Malaysia

(Chairperson)

Ain Nadzimah Abdullah, PhD

Associate Professor

Faculty of Modern Languages and Communication

Universiti Putra Malaysia

(Member)

Helen Tan, PhD

Lecturer

Faculty of Modern Languages and Communication

Universiti Putra Malaysia

(Member)

BUJANG BIN KIM HUAT, PhD

Professor and Dean

School of Graduate Studies

Universiti Putra Malaysia

Date:

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: _____

Name of
Chairman of
Supervisory
Committee: _____

Signature: _____

Name of
Member of
Supervisory
Committee: _____

Signature: _____

Name of
Member of
Supervisory
Committee: _____

TABLE OF CONTENTS

ABSTRACT	Page i
ABSTRAK	iii
ACKNOWLEDGEMENTS	v
APPROVAL	vi
DECLARATION	viii
LIST OF TABLES	xii
LIST OF APPENDICES	xv
LIST OF ABBREVIATIONS	xvi
CHAPTER	
1 INTRODUCTION	1
1.1 Background of the study	1
1.1.1 Disciplinary community	2
1.1.2 Academic genre	3
1.1.3 Corpus-based research and academic discourse	4
1.1.4 University lecture as a spoken academic genre	5
1.1.5 Lexical bundles and academic lecture	7
1.2 Statement of the problem	8
1.3 Objectives of the study	9
1.4 Research questions	9
1.5 Theoretical framework of the study	10
1.5.1 Linguistic competence and formulaic language	11
1.5.2 Concept of lexical bundle	15
1.6 Significance of the study	18
1.7 Definition of key terms	19
2 LITERATURE REVIEW	21
2.1 Introduction	21
2.2 Lexical bundles and written academic discourse	21
2.2.1 Lexical bundles and disciplinary study	21
2.2.2 Lexical bundles and cross-disciplinary study	23
2.2.3 Lexical bundles and cross-linguistic study	28
2.3 Lexical bundles and spoken academic discourse	32
2.4 Summary	36
3 METHODOLOGY	38
3.1 Introduction	38
3.2 Conceptual framework	38
3.3 Research design	39
3.4 Establishing the corpus	39

3.5 Sampling procedure	40
3.6 Framework of Analysis	40
3.6.1 Lexical bundle identification	41
3.6.2 Analysis procedure	42
3.6.3 Classificatory system of lexical bundles	43
3.7 Summary	48
4 RESULT AND DISCUSSION	49
4.1 Introduction	49
4.2 Lexical bundles in academic lectures	50
4.2.1 Lexical bundles in academic lectures of social sciences	51
4.2.2 Lexical bundles in academic lectures of physical sciences	52
4.2.3 Lexical bundles in academic lectures of life sciences	54
4.2.4 Comparison between the frequency of the bundles in the three corpora	55
4.2.5 Patterns and variations	56
4.3 Structural analysis of lexical bundles in academic lectures	60
4.3.1 Structural distribution of lexical bundles in SS	60
4.3.2 Structural distribution of lexical bundles in PS	62
4.3.3 Structural distribution of lexical bundles in LS	64
4.3.4 The comparison between structures of lexical bundles in the three Corpora	66
4.4 Revision of the taxonomy proposed for the functional classification of lexical bundles	77
4.5 Functional analysis of lexical bundles in academic lectures	79
4.5.1 Functional classification of lexical bundles in SS	79
4.5.2 Functional classification of lexical bundles in PS	82
4.5.3 Functional classification of lexical bundles in LS	86
4.5.4 The comparison between the functions of lexical bundles in the three corpora	89
4.6 Summary	123
5 CONCLUSION	124
5.1 Summary of key findings	124
5.2 Limitations and suggestions for pedagogical applications	128
5.3 Implications	129
REFERENCES	130
APPENDICES	139
BIODATA OF STUDENT	186
LIST OF PUBLICATIONS	187

LIST OF TABLES

Table	Page
3.1. Details of the corpus used in the study	40
4.1. The most frequently used lexical bundles in the corpus	50
4.2. Description of the SS corpus	51
4.3. Frequency information on SS lexical bundles	51
4.4. The top ten most frequent bundles in SS	52
4.5. Description of the PS corpus	52
4.6. Frequency information on PS lexical bundles	53
4.7. The top ten most frequent bundles in PS	53
4.8. Description of the LS corpus	54
4.9. Frequency information on LS lexical bundles	54
4.10. The top ten most frequent bundles in LS	55
4.11. Frequency information of lexical bundles in the three corpora	55
4.12. Most frequent 50 four-word lexical bundles in the three sciences	58
4.13. The top 5 bundles specific to each corpus	59
4.14. Distribution of lexical bundles across structural types in SS	61
4.15. Distribution of lexical bundles across structural types in PS	63
4.16. Distribution of lexical bundles across structural types in LS	65
4.17. Distribution of lexical bundles across the main structural types in the three corpora	67
4.18. Distribution of verb phrase bundles across the three corpora	69
4.19. Distribution of dependent clause bundles across the three corpora	73
4.20. Distribution of noun and prepositional phrase bundles across the three corpora	75
4.21. Functions of lexical bundles in academic lectures	78

4.22.	Frequency of lexical bundles across functional categories in SS	79
4.23.	Functional distribution of lexical bundles in SS	80
4.24.	Frequency of lexical bundles across functional categories in PS	82
4.25.	Functional distribution of lexical bundles in PS	83
4.26.	Frequency of lexical bundles across functional categories in LS	86
4.27.	Functional distribution of lexical bundles in LS	87
4.28.	Frequency of lexical bundles across functional categories in the three corpora	90
4.29.	Epistemic stance bundles in the three corpora	92
4.30.	Desire bundles in the three corpora	94
4.31.	Obligation/directive bundles in the three corpora	96
4.32.	Intention/prediction bundles in the three corpora	97
4.33.	Ability bundles in the three corpora	100
4.34.	Proportional distribution of stance bundles in the three corpora	101
4.35.	Topic introduction/focus bundles in the three corpora	102
4.36.	Topic elaboration/clarification bundles in the three corpora	105
4.37.	Proportional distribution of discourse organizing bundles in the three corpora	106
4.38.	Identification/focus bundles in the three corpora	108
4.39.	Imprecision bundles in the three corpora	109
4.40.	Quantity specification bundles in the three corpora	110
4.41.	Tangible framing bundles in the three corpora	112
4.42.	Intangible framing bundles in the three corpora	113
4.43.	Place reference bundles in the three corpora	114
4.44.	Time reference bundles in the three corpora	115
4.45.	Multi-functional reference bundles in the three corpora	116

4.46.	Proportional distribution of referential bundles in the three corpora	117
4.47.	Retrospective bundles in the three corpora	118
4.48.	Prospective bundles in the three corpora	120
4.49.	Interactional bundles in the three corpora	121
4.50.	Proportional distribution of lecture-specific bundles in the three corpora	123
5.1.	The top 5 frequent lexical bundles in the three sciences	125



LIST OF APPENDICES

Appendix	Page
1. Lexical bundles which were excluded from the SS corpus	140
2. Target bundles in the SS corpus	142
3. Lexical bundles which were excluded from the PS corpus	148
4. Target bundles in the PS corpus	149
5. Lexical bundles which were excluded from the LS corpus	158
6. Target bundles in the LS corpus	160
7. Shared bundles between the three sciences	165
8. Shared bundles between PS and LS	167
9. Shared bundles between PS and SS	170
10. Shared bundles between SS and LS	173
11. Lexical bundles specific to the SS corpus	176
12. Lexical bundles specific to the PS corpus	179
13. Lexical bundles specific to the LS corpus	184

LIST OF ABBRIVIATIONS

SS	Social Sciences
PS	Physical Sciences
LS	Life Sciences
VP	Verb Phrase
NP	Noun Phrase
PP	Prepositional Phrase



CHAPTER ONE

INTRODUCTION

This chapter starts by presenting some background information regarding the notion of discourse community, disciplinary community, genre and academic university lecture as a spoken genre. Next, statement of the problem, objectives of the study, research questions and significant of the study are presented. This is followed by providing information concerning the theoretical framework of the study, (communicative competence) followed by other theoretical issues that define characteristics of language knowledge that locate the importance of linguistic and pragmatic knowledge that concerns the use of lexis. Then the discussion moves to the concept of formulaic language as a form of the lexicon, and in its narrow sense, lexical bundles as the focus for this investigation. Finally, definition of key terms is presented.

1.1 Background of the study

The ultimate purpose of a language is to communicate knowledge and information among groups of individuals (Borg, 2003). Each of these groups is what Swales (1990) refers to as discourse community. He describes discourse communities as groups of members who share common goals and interests. In fact, they are the defining characteristics of discourse communities. The notion of discourse community is developed from speech community which refers to a group of people with shared functional rules and who recognize their language to be different from other languages (Paltridge, 2006). According to Hymes (1972), a speech community is a community “sharing knowledge of rules for the conduct and interpretation of speech. Such sharing comprises knowledge of at least one form of speech, and knowledge also of its patterns of use. Both conditions are necessary” (p. 51).

There are a number of differences between the concepts of discourse community and speech community. The first difference lies in the mode of communication. It is rather obvious that speech community is primarily characterized by speech, while a discourse community takes on a form of writing. In addition, discourse community is not necessarily constrained to specific space and time (Swales, 1988). Another difference lies in the way how members are regarded. Speech communities usually accept their members accidentally, while discourse communities tend to separate their members according to their specialties or interests. In his influential book *Genre Analysis*, Swales (1990, pp. 24-27) suggested six defining characteristics for identifying a group of individuals as a discourse community: First, a group of individuals must have some commonly shared set of public goals to be grouped in a discourse community. These public goals are supposed to create a communality of interest among the members of the communities like those of sporting clubs or associations. Second, a discourse community has mechanisms for the communication among its members. In Herrington's (1985) words, it has ‘a forum’. There are different kinds of mechanisms through which communication can take place, such as meetings, tele messages, newsletters, conversations, bulletins, etc. The third

characteristic is summarized into the fact that in each discourse community, information and feedback are two outputs of these shared mechanisms. The purpose of exchanging information depends heavily on a set of common purposes, such as improving performance in an orchestra or in a football squad, making money in a brokerage house, or denting the research front in an academic department. Besides, discourse community uses one or more genres to fulfill its communicative aims. To Martin (1985, p. 250) “genres are how things get done, when language is used to accomplish them”. Another characteristic emphasizes that apart from its generic features, a discourse community has some specific public lexicon. This is more apparent in the development of community-specific short forms and abbreviations to represent the jargon, such as EAP or ESP. Understanding these abbreviations are hard for people who are outside these discourse communities. Finally, each discourse community has a critical group of members familiar with a high degree of relevant content and discursial expertise. Membership in a discourse community is subject to change, with members entering as a novice individual and leaving by death or other involuntary ways.

Swales’s notion of discourse community has been a topic of contention in spite of its popularity. Two main arguments have been proposed by Duszak, (2005). First, the definition of discourse community is too narrow and does not contain an evolving nature. Second, the social approach proposed by Swales (1990) in relation to discourse communities “fails to account for the dynamics of how the core and the periphery co-construct each other” (Duszak, 2005, p. 70). In spite of the fact that the social approach introduced by Swales (1990) put emphasis on the essential characteristics of the text and its function, it still lacks to take into account the identities of human and institutions (Connor, 1996; Grabe & Kaplan, 1996; Ivanič, 1998).

1.1.1 Disciplinary community

Many of these discourse communities focus on particular domain of academic knowledge related to specific research and theory which are referred to as disciplines (Becher, 1994). A discipline is the knowledge of specific subject area shared by the members of the same discourse community. In other words, every discourse community is made up of a number of disciplinary communities which are identified with particular academic knowledge (Swales, 1990). In addition, scholars have broadened another notion of community which they term as disciplinary community (Bailey, 1977; Becher, 1994; Kuhn, 1970). This disciplinary community constructs social knowledge through a sense of agreement among the members of communities rather than individuals in a discourse community. As defined earlier, these groups of people interact with one another and share common ideas to form discourse communities (Swales, 1990). Becher (1994, p. 153) defines disciplinary communities as “academic tribes, each with their own set of intellectual values and their own path of cognitive territory”. Kuhn (1970) also considers a disciplinary community as “a group of practitioners with shared literature, judgment, communicative networks and professional goals” (p. 177). Bailey (1977) looks at disciplinary communities as cultural “tribes” and asserts that:

Each tribe has a name and a territory, settles its own affairs, goes to war with others, has a distinct language or at least a distinct dialect a variety of symbolic ways of demonstrating its apartness from others. Nevertheless the whole set of tribes possess a common culture: their ways of constructing the world and the people who live in it are sufficiently similar from them to be able to understand, more or less, each other's culture and even, when necessary, to communicate with members of other tribes (p. 212).

Following these definitions, several researchers have tried to categorize disciplines according to specific knowledge. Biglan (1973) categorized academic disciplines which are based on their subject matters into two main divisions; hard-soft and pure-applied. A field like philosophy is soft, showing characteristics oriented towards being personal and idiosyncratic. Hard fields like physics display the writer or speaker's intention towards a degree of agreement. Applied fields like mechanical engineering deal with practical and functional problems, as opposed to pure fields like history. Kolb (1981) categorized disciplines in a different way by focusing on four models of learning: concrete, abstract, active, and reflective, but perhaps the most known work in the area of disciplinary categorization is that of Becher (1989, 1994). He divides academic disciplines into four different categories of humanities and pure social sciences (soft pure), natural sciences (hard pure), applied social sciences (soft applied), and science-based professions (hard applied).

1.1.2 Academic genre

Disciplines located within specific knowledge systems give rise to another dimension of discourse classification. The term genre was established to give discourse a more sensitive description that highlights specific characteristics pertaining to the particular genre. The last 50 years has witnessed a growing body of research in the domain of genre as evident in the studies by Barber (1962), Herbert (1965), and Ewer and Hughes-Davies (1971). Swales (1990) develops a more comprehensive and salient theory for genre analysis with the main focus on academic genre. He defines genre as "a class of communicative events, whose members share some set of communicative purposes" (Swales, 1990, p. 58). It is generally believed that genres are characterized through communicative goals achieved by the members. Such communicative goals are said to constitute building blocks of academic genres (Swales, 1990). Bhatia (1993, p. 13) addresses the social aspects of these purposes, asserting that these communicative purposes only exist "within the framework of socially recognized purpose(s)". Bex (1996) also looks at genre from a social point of view and defines genre as "aggregation of communicative events that fulfill a common social function" (p. 137).

In 2004, Swales attempted to present a clearer picture of genre than the earlier one in 1990. Later he explains that "genres are seen metaphorically as frames for social action, not as social action themselves, providing only a relatively small part of what might in the end be needed for fully effective communicative action" (Swales, 2004, p. 61). Academically, genres are divided into spoken and written genres. An example of written genre is the research article, while the university lecture which is the focus of the present research is an example of the spoken genre. Several researchers have pointed out the typical differences between the spoken and written genre. Halliday (1985) proposes a number of differences between speech and writing. First, we speak

in order to perform actions, and this indicates the dynamic nature of speech. Writing, on the other hand, does not emphasize events. Second, speech deals with dialogues, while writing has to do with monologues. Third, speech comprehension is largely characterized by the instant context of situation, while it is the text which composes the written context. Finally, speaking is unpredictable, whereas writing is characterized by a systematic pattern.

Biber (1988) looks at the variations between spoken and written genres from the lexico-grammatical point of view. He claims that the written mode is characterized by prepositions, agentless passives, nominalizations, and low type/token ratio, whereas spoken mode adheres to a high dominance of first and second person pronouns, *that*-clauses, subordinate and conditional clauses, and contractions. He generally considers a spoken language as restricted, informal, involved and contextualized (Biber, 1988).

In line with the variations between the two types of discourse proposed by Halliday and Biber, Crystal (2004, p. 291) offers other differences between the spoken and written modes. First, while writing is more space-bound, static and permanent, speech is more time-bound, dynamic and transient. Second, unlike writing, speech does not include any time lag between the production and reception processes. Due to the lack of time, there is always a pressure to think while speaking, and this would lead to less repetition, rephrasing and comments. Another difference is that speech relies heavily on extralinguistic signs such as gestures and facial expressions to best express their meaning, while in writing, direct reaction is not in the same fashion. The next distinction implies that speech contains different examples of informal language such as slang which does not appear in writing. While in speaking, errors cannot be corrected, writing errors however, can be simply edited in the final draft. Crystal (2004) asserts that the two modes are also different in terms of their exclusive characteristics. Intonation, sound, tempo, rhythm and other tones of voice are characterizing features of speech which cannot be written down. On the other hand, pages, lines, capitalization and many examples of punctuation which cannot be read aloud are considered as features of writing. Given the distinctions, it is clear that each genre deserves attention on its own especially as a field of research. In genre research, it has become increasingly popular to investigate the associated discourse through large corpora that are established to capture language in use.

1.1.3 Corpus-based research and academic discourse

Before the advances in computer sciences, any researcher who intended to analyze the samples of spoken or written language had to provide and compile hard copies of all the materials and analyze them manually (Sinclair, 1991; Wray, 2002). This furnished the researchers with some qualitative realizations about the language use in the corpus, but the method could not provide a systematic quantitative overview of the language use. In the 20th century, advances in computer technology have provided researchers with a unique window to have access to an abundance of transcribed texts containing millions of words, thus giving a chance to store them electronically for macro and micro analysis of language use.

In recent years, the number of studies with specialized corpus to investigate the language use of different genres or registers has been increasing. Many of these studies have focused on analyzing written registers since the process of corpus construction is easier than that of spoken registers. However, recently, some corpora such as British Academic Spoken English (BASE) and Michigan Corpus of Academic Spoken English (MICASE) have been specifically compiled to reflect the use of language in academic oral settings. The two corpora contained transcribed samples across a variety of range of spoken registers such as academic lectures, seminars, discussion sections, interviews, study groups, etc. Therefore, corpus-based research has allowed for the investigation of different linguistic features in academic oral speech of language users. It was not possible to investigate large stretches of discourse without the development of computer programmes. Oral speech can manifest itself in a number of sub-genres such as meetings, oral presentations, group discussions, interviews and oral lectures. In the case of the BASE corpus, a larger number of lectures have been transcribed and compiled, producing a large database for the investigation of different linguistic elements in language use of four broad disciplinary divisions. This study has resorted to this corpus to obtain the relevant data for investigation into the use of lexical bundles in academic lectures across disciplinary divisions.

1.1.4 University lecture as a spoken academic genre

In the spoken sub-genre, university lecture is representative of a form of spoken academic discourse. Its significance is seen in lectures being acclaimed as “one of the most important teaching methods in higher education institutions” (Thompson, 1994, p. 171). A lecture is simply defined as an extended piece of discourse that is presented by one speaker to a group of listeners. According to Flowerdew (1994), academic lectures serve as a specific pedagogical and instructional genre which contributes to students’ better perception of the subject materials and courses. Like any other academic genre, communicating the knowledge base of a discipline to the listener is also considered as a principal purpose of academic lectures. Since much of what learners acquire in class is through lectures, it is necessary to explore different aspects of this spoken sub-genre in terms of its features, such as form and function, in order to familiarize the learners with the oral characteristics of the language being used in academia. Despite its pervasiveness, learning this type of genre is not easy for many language learners who may encounter a variety of linguistic difficulties in the process, such as listening to and comprehending long lectures within the context of a particular discipline. Meanwhile, lecturers also find it difficult to present and organize the disciplinary materials in a way that is comprehensive for the learners.

Academic lectures serve different functions based on their purpose upon which the lecturer presents the materials. There is a common consensus among the researchers on some general functions for lectures, such as providing the learners with information that cannot be found in written materials such as textbooks (Young, 1990), making a logical connection between what has been presented in the previous lecture and what to be taught in the current lecture (Thompson, 1994), and using multiple examples to explain a complex idea (Young, 1994). A few scholars use the term “role” to describe lectures and three primarily roles have been identified. The first is passing on of information or what Barr (1990, p.6) referred to as “informing

role”. Second is the “evaluating role” (Dudley-Evans & Johns, 1981, p. 32) which entails what the lecturers evaluate from the subject matter, and third, “organization role” which shows the overall organization of the lecture (Thompson, 1994, p. 172). Concerning the last role and also in line with the overall organization and structure of academic lectures, Young (1990, 1994) identifies six phases that constitute the structure of lectures, namely discourse structuring, conclusion, evaluation, interaction, theory or content, and examples. The discourse structuring phase is significantly crucial since it assists the audiences in the listening event and directs them towards the flow of the lecture by using expressions like *Today, we’re going to talk about*, and also helps the listeners to make sense of new information (Lee, 2009).

Lecturing styles are also different based on a variety of elements such as the lecturers’ training, personality and experience (Nattinger & DeCarrico, 1992). Some lectures are more like conversations, with the frequent use of informal language or idiomatic expressions. Others follow the style of written materials largely using formal and prescribed language. For the most part, every lecture is believed to follow one of the styles. However, Nattinger and DeCarrico (1992) stress that some lectures may combine the two styles, with the lecturer initiating the lecture in a more formal way, and then gradually moves into a more relaxed and informal manner in the body section of the lecture by using more idiomatic language. Following the discussion, Dudley-Evans (1994, p. 148) offers three main styles of lecturing:

1. In *reading style*, the lecturer presents the lecture either from the notes prepared in advance or performs as if he is reading from notes, and such notes seem to be retrieved from memory. This style is normally characterized by little interaction between the lecturer and the students.
2. In *conversational style*, the lecturer delivers the lecture making a frequent use of informal language such as idiomatic expressions, and there is a notable amount of interaction with the students.
3. In *rhetorical style*, the lecturer acts as if he is a performer using a range of intonations.

From the discursal point of view, Nattinger and DeCarrico (1992) propose two different kinds of styles in determining the discourse of academic lectures, namely interactional and transactional. In general, most academic lectures comprise both transactional and interactional discourses; however, the interactional part is minimal (Nattinger & DeCarrico, 1992). Both interactional and transactional discourses are distributed across sections of the lectures, leading to the need to structurally examine different sections of the lectures, namely, introduction (or opening), body and closing. In her significant work on lecture introductions, Thompson (1994) asserts that what makes the introduction section of the lecture noteworthy is the opportunity they offer to the lecturers “to establish an interpretive framework for the audience to use as they listen to the rest of the lecture” (p. 174). The introduction normally serves as a point of departure in the lecture, starting with some interactional expressions such as greeting, short questions and answers, or talking about the weather. In some cases, the lecturer may use some topic marker expressions which have the dual function of announcing the end of the interactional part and laying the ground work to move to the transactional part and start the body section. The body section of the academic lectures is completely transitional, requiring the lecturer to explain the detailed content of the lesson or subject materials. This section is actually the general

scaffolding of the lecture which the content fits into (Nattinger & DeCarrico, 1992) and thus can be considered as the section with the most potential to investigate a variety of linguistic features. The closing section of academic lectures is the shortest, yet it is as important as the body and the opening sections because “a strong and powerful ending often stays clearly in the listeners’ minds” (Cheng, 2012, p. 234). The ending mostly includes interactional expressions of farewell or leave-taking exchanges between the lecturer and the students. In some cases, the closing section has a transactional phase, with the lecturer briefly introducing the topic of the next lecture. Besides all these features, the style of delivery obviously will vary across different disciplines and sciences.

1.1.5 Lexical bundles and academic lecture

In research on spoken registers such as academic lectures, abundant attention has been given to analyzing lexical phrases (De Carrico & Nattinger, 1988). Lexical phrases are considered as language clusters which vary in length such as *in terms of*, *on the other hand*, *it is necessary to mention* and there has been a realization that mastery of these language clusters could ease the difficulties learners have in listening comprehension. In recent years, with the advances in technology and the emergence of new methodologies for corpus-based research, the analysis of these clusters in oral registers has been done on a higher scale.

In order to give more background to the study, in this section, a type of lexical phrase which helps to organize the discourse of academic lectures is introduced. One type of lexical phrase can be in the form of formulaic expressions. To begin with, in the delivery of academic information in university lectures coming from different sciences, the lecturer relies on the deployment of sets of common formulaic expressions in order to present the disciplinary materials in a more coherent way. One type of formulaic expressions is what Biber et al. (1999) called lexical bundles. They define lexical bundles as a combination of more than two words which co-occur frequently in a given register, helps to organize the discourse. The common examples which are used in academic language are *on the other hand*, *in terms of the*, *as can be seen*.

The four-word string is the most common and researched length for lexical bundles because the “number of four-word bundles is often within a manageable size (around 100) for manual categorization and concordance checks” (Chen & Baker, 2010, p. 32). They are more recurrent than five-word bundles in academic studies and mostly hold three-word bundles in themselves. For example, the four-word bundle *in terms of the* already includes the three-word bundle *in terms of*. Some lexical bundles are not opaque in terms of their semantic and their communicative function may vary from one discipline to another discipline or from one register to another register. Bundles are common, transparent sequences which co-occur with a high frequency within a discourse community (Biber et al., 2004). In identifying lexical bundles, frequency of occurrence and distribution parameters are the key features. To identify that a string of word operates as a lexical bundle in a text or speech, a researcher uses computer applications to identify them and then proceeds to decide the frequencies and also to analyze the environment or context of their occurrences. More information on this core concept of lexical bundles is given in chapter 2.

1.2 Statement of the problem

When students have to contend with a system of oral content delivery, it is widely realized that listening to lectures could put language learners in difficulty both linguistically and cognitively (Flowerdew, 1994; Flowerdew & Miller, 1997; Thompson, 2003). Comprehending lecture monologues and finding the best relationship between the word elements used have always been a major problem for learners as they make sense of the message. In particular, non-native speakers may find the task extremely challenging. They have to deal with content delivered through a variety of language features which may be alien to them. Among them could be the deciphering of underlying meaning conveyed by formulaic combinations. Even proficient learners who may make sense of the meaning of individual words of a lecture, may have problems in recognizing the logical relationship between the word elements. This would undoubtedly result in imperfect comprehension of the lecture. A number of reasons could be suggested for such a phenomenon. Perhaps the major reason could be traced to the real-time processing feature of listening comprehension which distinguishes it from other types of comprehension such as reading comprehension. Unlike readers, listeners of academic lectures do not normally have the same amount of time to process the information, since the utterance and the understanding of the lecture must occur simultaneously. In addition, academic lectures are transient, in that, once the lecturer utters the words, they disappear. The audiences cannot return exactly word for word to what the lecturer had presented unless they have recorded them. This is what makes listening more challenging than reading.

Another key element is lexical ignorance, which seems to be one of the main hindrances to academic lecture comprehension, even with advanced learners. Lexical ignorance can be linked to the lack of familiarity with formulaic expressions which could be considered as a difficult feature in language use. In the context of academic oral discourse, such expressions similarly could pose problems to language users. Students may not be familiar with the structural and discursive functional characteristics of formulaic sequences, such as lexical bundles as they have found to serve more than one function based on discipline or register. In the course of learning a language, many learners may consciously avoid more complex structures in their writing or speech. As such, there may be little awareness or limited contact with multi-word expressions that are more complex in their structure and function. ESL lecturers may also suffer the same limitations in language use thus resulting in ineffective delivery of the content in their disciplines.

In the context of these issues, what has motivated this study is the need to understand how oral discourse is structured with the use of lexical bundles. It is contended that understanding how lexical bundles are used will facilitate the oral comprehension as well as oral delivery especially for those who are entry ESL learners in a learning environment when English is dominantly used. As mentioned, initial language problems including, lexical ignorance could lead to further linguistic and cognitive difficulties as learners progress through the stages of learning. Thus, this study takes on the task of understanding the use of lexical bundles that would facilitate learning. In addition, lexical bundles are situated in specific discourse types or disciplines. They have shown to have communicative and discourse functions specific to

disciplines and sciences. This variation in oral speech is also attended to as part of the study so as to enrich the body of knowledge regarding the use of lexical bundles in general and specifically the data obtained could lead to greater awareness of how oral lectures are delivered according to discipline.

1.3 Objectives of the study

The focus of this study is on multi-word expressions or lexical bundles found in oral discourse represented by academic lectures. In the last two decades, the notion of lexical bundle has been widely investigated as a phraseological unit with the aim of comparing corpora of language in different registers, such as conversation and academic prose (Biber et al., 2004). Several other studies focused on the notion of disciplinary variations (Cortes, 2004; Hyland, 2008), particularly in written academic genres. However, little attention has been given to the use of four-word lexical bundles in spoken discourse, especially academic lectures of different disciplinary fields.

One of the effective ways for speakers or lecture presenters to form meaning and create a voice to make a good impression on the listener is through the use of formulaic expressions like lexical bundles. Therefore, students as well as lecturers coming from different disciplinary communities need to be aware of the importance of using lexical bundles in oral delivery of their specific disciplinary content. With this concern in mind, the present study aims to explore the use of frequently occurring four-word lexical bundles in academic lectures of three broad disciplinary divisions: life sciences, physical sciences, and social sciences so as to come up with empirical data on the possible similarities and variations regarding the frequency, structural characteristics and the discourse functions of the identified bundles. The three sciences were selected on the basis of fulfilling the purpose of the study which is to explore the bundle use in disciplinary “sciences” rather than non-scientific disciplines which would warrant another focus on its own.

1.4 Research Questions

The current study attempts to answer the following questions:

1. What are the most frequent four-word lexical bundles found in academic lectures of social sciences, physical sciences and life sciences?
2. How are the academic lectures in social, physical and life sciences different or similar in terms of frequency of occurrence of the lexical bundles used?
3. What are the structural characteristics of the lexical bundles used in the lectures of social, physical and life sciences?
4. How are the academic lectures in social, physical and life sciences different or similar in terms of the structure of the lexical bundles used?

5. What are the functional characteristics of the lexical bundles used in the lectures of social, physical and life sciences?
6. How are the academic lectures in social, physical and life sciences different or similar in terms of the discourse function of the lexical bundles used?

1.5 Theoretical framework of the study

Communicative competence is one of the theories of linguistics that accounts for the grammatical, sociolinguistic, strategic, and discourse knowledge of language users. The notion of communicative competence was coined by Hymes (1967; 1972) as a superior model of language and a reaction against Chomsky's (1965) "linguistic competence" which differentiates performance from competence. Hymes brought the sociolinguistic perspective of language into Chomsky's view of linguistic competence by focusing on the relationship between communicative form and function. Since then, a number of scholars have adopted this opposing view as a basis for describing the development of communicative competence. Canale and Swain (1980), and later Canale (1983), proposed the first comprehensive model of communicative competence, with the aim of serving both instructional and assessment purposes (Celce-Murcia, Dörnyei, & Thurrell, 1995). The model included four major components of communicative competence:

1. Grammatical competence, including the knowledge of vocabulary, pronunciation, grammatical rules, etc.
2. Sociolinguistic competence, including the appropriate use of the language in a given context.
3. Strategic competence, including the use of communication strategies.
4. Discourse competence, including the knowledge of language structure to form coherence and cohesion.

Bachman (1990) and later Bachman and Palmer (1996) decided to elaborate on the Canale and Swain's model by recasting the construct of communicative competence through the context of language testing which very much also reflects learning. Their model included two main categories for the knowledge of language, organizational knowledge and pragmatic knowledge. The organizational knowledge of language includes grammatical and textual knowledge, while pragmatic knowledge encompasses lexical, functional, and sociolinguistic knowledge.

Since Bachman (1990) and Bachman and Palmer's (1996) model has been designed with reference to language assessment rather than language teaching, a need was felt to look at communicative competence and their sub-components from pedagogical perspectives. To this aim, Celce-Murcia, Dörnyei, and Thurrell (1995) proposed the most recent model of communicative competence to introduce overt and potential principles to the instruction of communicative skills. They attempted to elaborate on the earlier models by adding a new competence, referred to as actional competence, to Canale and Swain's model, as a way to best realize the communicative purposes

by performing speech acts. Canale (1983) earlier derived discourse competence from sociocultural competence, and the model by Celce-Murcia, Dörnyei, and Thurrell (1995), even narrowed down the sociocultural competence by branching into actional competence. Another modification included changing the terms sociolinguistic competence and grammatical competence in Canale and Swain's model to sociocultural competence and linguistic competence, respectively. The latter modification was motivated by the need "to indicate unambiguously that this component also includes lexis and phonology in addition to morphology and syntax" (Celce-Murcia, Dörnyei, & Thurrell, 1995, p.11).

1.5.1 Linguistic competence and formulaic language

Celce-Murcia, Dörnyei and Thurrell (1995) introduced some basic elements of communication under the linguistic competence, including patterns and types of sentence, constituent structures, morphology, systems of phonology and orthography, and lexical knowledge (Figure 1). The importance of lexico-grammatical building blocks or what Pawley and Syder (1983) referred to as "lexicalized sentence stems" or "formulaic constructions" were highly emphasized in Celce-Murcia, Dörnyei and Thurrell's (1995) model. They believed that the knowledge of lexicon belongs to more than one area. The functional role of lexical phrases would fall under the category of discourse competence, while their systematic aspects, including their meaning and process of word-building belong to linguistic competence. Under the category of lexical knowledge, they proposed four sub-components, namely words, routines, collocations and idioms. Routine expressions include those word-like fixed phrases and formulaic and semi-formulaic chunks which are the focus of the present study (they are highlighted in Figure 1).

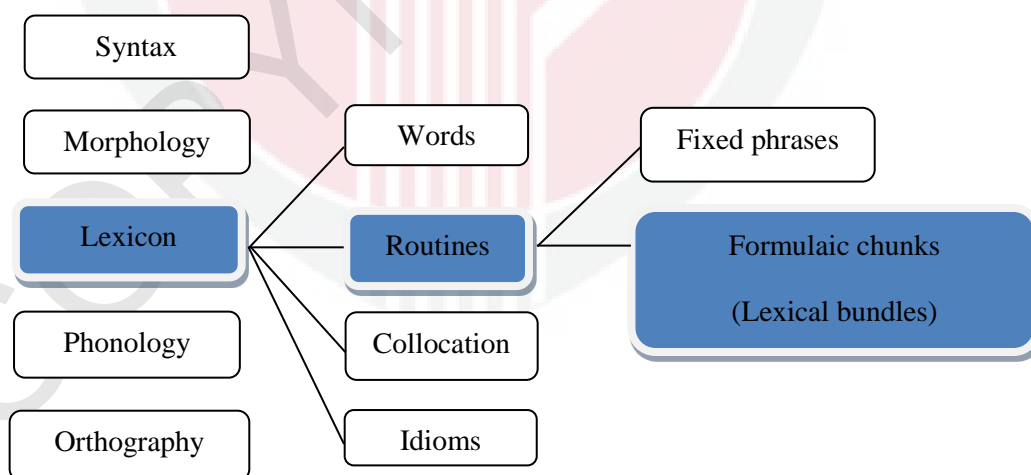


Figure 1. Linguistic competence and formulaic language (Celce-Murcia, Dörnyei & Thurrell, 1995)

In addition, the system of language has two primary processing styles; the generative rule-based view and exemplar-based view (Peters, 1983; Weinert, 1995). The rule-based view considers learning as a process of generating and transferring fundamental rules, which are taken from stimulus materials (Reber, 1989; Skehan, 1998). Supporters of the exemplar-based view believe that learning involves a

collection of formulaic sequences accumulated in a learner's memory, which are retrieved completely at the time of use (Bolinger, 1976; Lewis, 1993). These formulaic sequences are derived from linguistic competence and are believed to play a crucial role in fluent linguistic production of discourse, especially oral discourse (Biber et al., 1999, 2004; DeCock, 2004; Pawley & Syder, 1983; Schmitt, 2004; Wray, 2002, 2008). Natural language seems to be characterized heavily by remarkable use of these formulaic chunks of language. One can never use a language without coming across formulaic sequences, which are the way words come to co-occur with some neighboring words, such as *a lot of* or *for example*. Formulaic language is simply more than two independent morphemes which are stored in a speaker's lexicon as a single unit and they are not analyzed through grammatical rules. Wray (2002) defines a formulaic sequence as:

"a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar" (p. 9).

Schmitt and Carter (2004) also describe a formulaic sequence as a combination of words with a particular function that is automatically and frequently used by native speakers. It operationally refers to any word sequence, syntactically complete and semantically meaningful, which can be looked up in a reliable dictionary (Qi & Ding, 2011). Conklin and Schmitt (2008) argue that formulaic sequences are more than just chunks of word connected by collocational binds. Instead, they are dealing with communicative content of the language. Therefore, having a good command of these formulaic units would, needless to mention, bring naturalness, fluency and native-likeness to a learners' speech production (Conklin & Schmitt, 2008). In the same line, and to emphasize the effect of formulaic language on fluency, Wood (2001) postulates that speakers who have a good command of automatized and formulaic sequences of language are more able to "balance skills, attention and planning during speech" (p. 578).

In addition to contributing to fluency, formulaic language has been found to have several other communicative purposes related to the notion of pragmatics such as reducing the processing effort and ensuring the later acquisition and retention. In view of this, Wray (2000) regarded formulaic language as a time-buyer in oral interactions, trying to take and hold the turn while the speaker speaks (*as you know, let me just say*). In writing, formulaic language serves a discursive function and is used mostly as a marker to organize the language (*as can be seen, on the basis of*).

The study of formulaic language has a long history in applied linguistics and has attracted much researchers' and instructors' attention from the beginning of the last century. The literature dates back to Jespersen (1917, 1924), who introduced the term 'collocation' and fixed expressions, as a comparison to free expressions. Firth (1964) also used the terms "collocation" and "collocability" to refer to consistent occurrence of a word with other words. Following this, some new language theories and research methods regarding the acquisition and use of multi-word combinations were suggested, such as ethnography (Fillmore, 1979; Hakuta, 1974; Peters, 1983), conversational analysis (Manes & Wolfson, 1981; Tannen, 1987) and quantitative analysis of multi-word combinations (Altenberg, 1993; Kjellmer, 1991; Sinclair, 1991). They all aimed to replace traditional ideas about grammatical analysis which

looked at lexis and lexical choices as some limited slots made by grammar, instead of considering them as systematic structures with repetitive patterns in their use. In other words, “grammar is the output of repeated collocational groupings” (Hyland, 2008, p. 6). To Sinclair (1991), most of the text is made up of a number of frequent and common words which are components of a large stock of multi-word patterns of language which make the text.

It is widely realized that spoken language contains a greater range of formulaic sequences than written language (Pawley & Syder, 1983). It seems that since speaking allows the speaker to use language more quickly, it basically includes a larger portion of formulaic language than writing. On the other hand, in written registers, because the creation of manuscript requires more analytical methods, the language production contains lower range of pre-fabricated expressions and perhaps more “newly constructed segments” (Wood, 2001, p. 580). In addition, there is an assumption that time limits and speed of production are also considered as key factors which allow for a higher occurrence of formulaic language in speech than in writing (Kuiper, 1996).

Another piece of evidence that supports the claims about the importance of formulaic language in language production and fluency comes from psycholinguistic perspectives. It is generally assumed that while speaking fast, the speaker will rely more on the use of ready-made clusters and formulas to control his speaking. A study of this type by Kuiper (1996) confirmed that professional auctioneers and sports broadcasters employed a large proportion of prefabricated and fixed sequences. Although Pawley and Syder (1983) believe that a speaker cannot produce a clause of more than 8-10 words, “some clauses are entirely familiar, memorized sequences which the speaker or hearer is capable of consciously assembling or analyzing, but which on most occasions of use are recalled as wholes or as automatically chained strings” (p. 15)

Taking these argumentations, a main question arises, that is, how much of a person’s lexicon is made up of these formulaic units? Interest in using frequent word combinations and ready-made phrases like *have a nice day* or *see you tomorrow* increasingly has led researchers to suggest that around 80% of natural language is structured in this way (Altenberg, 1998). Several other studies have shown that almost one-third to one half of the language people are using is made up of formulaic components (Erman & Warren, 2000; Foster, 2001; Howarth, 1998). However, Biber et al., (1999; 2004) believe that this proportion can fluctuate depending on different modes and registers. Despite the general agreement on the pervasiveness and importance of formulaic sequence in language production and fluency, related empirical studies have little consensus on the way to identify them, and their characteristics, and even how to label them. Biber et al., (2004, p. 372) believe that these studies differ in terms of:

1. Their research purposes: exploring a large range of multi-word units in contrast to a small set of units.
2. The kind of criteria applied for the identification of multi-word combinations: whether this is perceptual salience, frequency point criteria, etc.

3. The formal features of the multi-word sequences under study: continuous sequences, discontinuous frames, or lexico-grammatical patterns; two word collocations or longer sequences.
4. The size of corpus under study. This ranges from a small number of texts in some studies to very large corpora (100 million words).
5. Whether register variations are taken into account or not. This notion is completely ignored in many studies; while some other studies focus on analyzing either spoken or written registers. A few others overtly compare the use of multi-word combinations in different registers.

Among an extensive number of quantitative studies on analyzing language corpora, the last three decades have witnessed two distinct principles to the identification of multi-word expressions. The first method was proposed by DeCarrico and Nattinger (1988) and Nattinger and DeCarrico (1992) based on the idea that particular chunks of words are considered more important because they are common, and thus, frequently used by native speakers of that language. It talks about the way that multi-word combinations are identified intuitively rather than empirically. To explore the collocations in a corpus of lectures, Nattinger and DeCarrico (1992) only considered lexical phrases as those strings of words which seemed to be prefabricated, rather than produced by syntactic competence (Nesi & Basturkmen, 2006). Therefore, the first step in the identification process following this method would be to be aware that a prefabricated chunk exists.

The second method which is the basis of the present study relies on a computer program that identifies frequency and length of words using a range of cut-off frequency norms (Altenberg, 1993; Biber et al., 1999; Butler, 1997). Unlike the previous method, multi-word units in this method are identified empirically rather than intuitively. Nesi and Basturkmen (2006, p. 25) stated that in this method, “strings of frequently co-occurring words can be identified within a given corpus regardless of syntactic boundaries or their salience as meaningful units independent of context.” Chunks of words identified by this method are labelled under different names, such as ‘clusters’ (Scott, 1997; Hyland, 2008), ‘recurrent word-combinations’ (Altenberg, 1998, p. 101), ‘prefabricated patterns’ (Granger, 1998), ‘statistical phrases’ (Strzalkowski, 1999), phrasal lexemes (Moon, 1998), ‘lexical bundles’ (Biber et al., 1999, p. 993), ‘formulaic sequences’ (Schmitt & Carter, 2004), and ‘n-grams’ (Banerjee & Pedersen, 2003), among others.

Research using this method was pursued by Altenberg (1993, 1998) who pioneered a methodology to recognize frequency-defined recurrent word combinations and categorized them in relation to grammatical and functional analysis. However, Biber et al. (1999) took the first significant step in the form of a corpus study which investigated multi-word combinations, by classifying their functional and structural types in written and spoken discourse. They referred to these newly found units as “lexical bundles.” Since then, advancements in computer programs designed for the analysis of language corpora and the increase of corpus linguistics studies have introduced some new paths for researchers to conduct more empirical studies to single out linguistic characteristics of different types of multi-word combinations, more specifically lexical bundles.

1.5.2 Concept of lexical bundle

In examining discourse from an idiomatic point of view, which looks at language comprised of sets of chunks and formulaic expressions, we should consider word-combination analysis. One way to accomplish this is by retrieving lexical bundles. In the last 20 years, many studies have stressed the importance of teaching and using multi-word units in a range of academic discourse (Biber, 2006, 2009; Biber & Conrad, 1999; Biber & Barbieri, 2007; Biber, Conrad & Cortes, 2004; Butler, 1997; Cortes, 2002, 2004; Cortes & Csomay, 2007, 2009; Hyland, 2008; Simpson-Vlach & Ellis, 2010; and Wray & Perkins, 2000).

With the advance in computer technology, the analysis and calculation of multi-word combinations have become much easier (Jablonkai, 2012). *The Longman Grammar of Spoken and Written English* (Biber, Johansson, Leech, Conrad & Finegan, 1999) was the first grammar published within a framework which focused on a special type of these word combinations called lexical bundles, also known as clusters or chunks (Scott, 1997; Hyland, 2008). A number of scholars have provided some operational definitions of lexical bundles in their related research (Biber et al., 1999; Biber & Conrad, 1999; Cortes, 2004; Hyland, 2008). In this regard, the series of comprehensive studies conducted by Biber and his colleagues (Biber et al., 1999; 2004; Biber & Barbieri, 2007; Biber & Conrad, 1999) on the use of lexical bundles in a variety of spoken and written registers are considered as pioneering studies for the understanding of the notion of lexical bundles and their characteristics and their work underlie the approaches taken in the present study.

In their first work on the use of lexical bundles in conversation and academic prose, Biber et al. (1999, p. 990) defined lexical bundles as “recurrent expressions, regardless of their idiomaticity, and regardless of their structural status,” kinds of word combinations that co-occur frequently and are basically identified empirically, rather than intuitively. The research was later expanded 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 conversation, classroom teaching and textbooks. Biber and Conrad (1999, p. 183) reworded the earlier definition and defined lexical bundles as “multi-word expressions which occur frequently and with accidental sequences of three or more words,” such as *in the case of the*, *as a result of*, and *on the other hand*, among many others. Lexical bundles are, in essence, frequently occurring chunks 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 the language (Biber et al., 1999; Biber & Conrad, 1999). Previous corpus-driven studies have also demonstrated the ubiquity of these sequences in a variety of academic genres (Hyland, 2008).

Biber et al. (2004) argued that frequency is a determining criterion for studying lexical bundles, as it is a “reflection of the extent to which a sequence of words is stored and used as a prefabricated chunk” (p. 376), with sequences occurring with very high frequency and across multiple texts are more likely to be stored in long-term memory as unanalyzed chunks. However, frequency is only one condition under which a sequence becomes prefabricated, for example, sequences containing idiomatic meanings are rarely used but undoubtedly prefabricated. This would lead to a notable consideration in view of the fact that lexical bundles appear to bridge

syntactic boundaries and are not idiomatic in meaning, and consequently, as Biber, Conrad and Cortes (2004) claim, not very salient:

for the most part linguists have not noticed these high frequency multi-word sequences, probably because most previous research has focused on grammatical phrases and clauses, disregarding the possibility of lexical units that cut across grammatical structures. (p. 377)

Another characteristic of lexical bundles is that they are not structurally complete. Biber et al. (1999) discovered that in conversation, only 15% of the bundles are made up of complete clauses or phrases, while this percentage decreases to five in academic prose. Bundles also have different structures in relation to spoken and written discourse. As for spoken discourse, such as in conversations, majority of the multi-word combinations are made up of clause, of the type (pronoun) + verb phrase, such as *you want me to* and *have a look at*, while in academic prose, 60% of the expressions are parts of noun phrases or prepositional phrases such as, *as a result of*, *on the basis of*, and *on the other hand* (pp. 993-1000).

Hyland (2008) built his definition on the idea that lexical bundles play an important role in building coherence in a text or understanding the meaning of particular contexts. For example, phrases like *what I want to say* or *as far as I know* refer to group discussion and conversation, while *it is noted before* or *in accordance with* belong to academic prose. Wray and Perkins (2000) also pointed out that lexical bundles serve as types of short-cuts which are stored in memory and retrieved later at the time of use rather than making a new sequence each time. Therefore, knowledge of lexical bundles can reduce the amount of time spent on looking for the same familiar sequence of words when it comes to practice.

The notion that frequency of occurrence is a vital factor in determining lexical bundles is key to a common definition in the literature. Cortes (2004) defined lexical bundles as “extended collocations, sequences of three or more words that statistically co-occur in a register” (p. 400). She proposed three main characteristics that distinguish lexical bundles from other types of word combinations such as idioms or collocations. They are: frequency, idiomaticity, and fixedness. It is realized that every string of words should meet these qualities to be called a lexical bundle. The first and the most important characteristic is frequency of occurrence (Altenberg, 1993; Biber et al., 1999; Biber & Conrad, 1999). A normal occurrence of a lexical bundle is more than 20 times per million words, but in certain registers, some bundles may occur over 100 times in a million words, depending on the context. This rate of frequency is not comparable with other forms of word combinations such as in the idiom *kick the bucket*, which the occurrence is only 0.5 times in a million words (Cortes, 2004). With the growth of corpus-based quantitative studies using specific computer programs to identify word combinations, analyzing the frequency of lexical bundles in different registers has become more possible.

The second characteristic is idiomaticity, which suggests the idea that some words make new meanings when they combine with other words or strings of words (Chafe, 1968; Lattey, 1986; Moon, 1998). However, unlike idioms, in a lexical bundle, the new meaning can be easily retrieved from the meaning of its individual words. Examples are *it should be noted that*, *on the basis of*, and *in relation to the*, compared with the idiom *kick the bucket*.

The last feature of bundles is fixedness. Fixedness refers to the limitation in the occurrence of word combinations (Aisenstadt, 1981; Allerton, 1984; Bolinger, 1976). Lexical bundles are structurally and syntactically fixed; in that, you cannot find other forms of the same bundle in the text. For example, in many academic texts, only the singular structure *on the other hand* is considered as a lexical bundle. The plural expression *on the other hands* is not used as a frequently recurrent phrase. Cortes (2004) refers to this type of expressions as ‘frozen fixed.’ However, some other bundles allow slight modifications, depending on the context, such as *be in a public* which is “marked for tense and aspect” (p. 400).

In order to identify lexical bundles in a text or speech, two different criteria are required. The first is to set a cut-off frequency point. Since frequency is the determining feature in qualifying lexical bundles, researchers should be on their toes in defining the identification parameters. In general, there is a shared agreement that the definite frequency cut-off used to identify a lexical bundle is something arbitrary, depending on the corpus and number of texts (Adel & Erman, 2012; Biber & Barbieri, 2007; Biber & Conrad, 1999; Biber et al., 2004; Cortes, 2004; Hyland, 2008). Studies using a small corpus normally practice a lower cut-off point such as 10 times per million words, while those using larger corpora may use frequency rate of 40 times per million words (Biber, 2006). But typically for a study with a corpus of around 1000,000 words (such as the present study), it is normal to set the cut-off point at 20 times per million words (Chen, 2008; Cortes, 2004, 2006, 2008; Hyland, 2008). However, it is prevalent for many lexical bundles to occur much more frequently than this rate in a corpus, as much as 200 times per million words (Biber, 2006). However, a word of caution is needed in following this method. There is one disadvantage regarding this method of identification, which is, it does not identify expressions with discontinuous frames (*not only ... but also*), as Nesi and Basturkmen (2006) explain, longer sequences are not identified due to the pre-specification of length of strings.

While frequency is considered as a determining factor in defining bundles, it does not solely show the formulaicity of the bundle. Biber, Conrad and Cortes (2004) state that “frequency is only one measure of the extent to which a multi-word sequence is prefabricated” (p. 376). Therefore, in addition to frequency, another parameter which is dispersion or distributional pattern is also required to ensure that a string selected for analysis is not used in the given register by chance. Normally, most studies use a criterion of occurring in at least 5 different texts. It is necessary to mention that, here by text, we mean every kind of discourse in the form of transcript such as lecture, research article or even book chapter. This criterion is used to guard against individual writer or speakers’ idiosyncratic influences (Biber, 2006).

Even though, they are structurally incomplete, lexical bundles are “important building blocks in discourse” (Biber & Barbieri, 2007, p. 270). There is a well-established assumption that since lexical bundles are common, they can be easily acquired in a natural process of language learning. However, Biber and Conrad (1999) have argued that regardless of their frequency, lexical bundles are complex and not fixed. Second language learners have long realized that frequent use of such formulaic sequences would empower their speaking fluency and make their speech sound native-like, as Hyland (2008) believes; these multi-word clusters are the main

source of fluent linguistic production. These qualities emphasize the need to continue exploring lexical bundles in the language use of different discourse communities.

1.6 Significance of the study

Research in the area of formulaic language is of a great value in applied linguistics. A realization of how different word elements come together and create a new string of words with a different function can create a good opportunity to encourage and teach language learners from different disciplines to use them in their speech or writing. In the last two decades, there has been a great deal of interest in conducting studies on formulaic expressions. These refer to those frequently occurring sets of words that are used as a single item in a sentence and play a crucial role in defining membership in disciplinary communities. As Wray (2002) argues, formulaic sequences can determine language users' identity in communicating with particular groups such as discourse communities. Every disciplinary community shares a set of public goals which are common among its members and they have to acquire the necessary language skills to communicate efficiently.

One of the best ways to communicate effectively is to have good knowledge of formulaicity evident in different types of recurrent word combinations. An example of these frequently occurring sets of words is lexical bundles or chunks or clusters which are in essence, "words which follow each other more frequently than expected by chance" (Hyland, 2008, p. 5). They help to shape meaning in both written and spoken discourse. A number of researchers believe that lexical bundles are important to language performance because they account as a significant feature of the way we understand and also the way we speak (Bolinger, 1975; Peters, 1983). In addition, these bundles also play an important role in specifying particular register or discourse. For example, bundles like *as a result of* and *as can be seen* are more likely to refer to an academic written register, whereas, *as far as I know* and *what I want to say* seem to belong to a spoken register.

Haswell (1991) emphasizes that lacking the knowledge of lexical bundles would account for the characteristic behavior of a novice speaker or writer. There is no doubt that fluent speakers or writers rely extensively on the use of such formulaic sequences, because applying these handy expressions helps learners to communicate better and also the ability to get their meaning will save cognitive processing time in understanding language use. On a similar note, Coxhead and Byrd (2007) offer three reasons for the importance of lexical bundles to writers and speakers:

- (1) The repetitive nature of lexical bundles provides users, more specifically students, with some prepared strings of words to work with.
- (2) They contribute to fluent speech production and as a result establish disciplinary membership.
- (3) They show some lexico-grammatical and community-authorized methods to make meaning.

In addition, having a good command of lexical bundle use can to a great extent, ease comprehension problems language learners could encounter while listening to academic lectures in their disciplines. In academic discourse, lexical bundles function as creating logical relationships between the ideas in sentences (*in contrast to the, is due to the*), guiding listeners or readers through a lecture or text (*as mentioned earlier, in the next session*) and building coherence in speech and writing. Therefore, findings from this study could feed invaluable information into the formulation of teaching methodologies or EAP training approaches in terms of strategy training. Furthermore, knowing how the understanding of academic lectures can be facilitated will assist disciplinary lecturers working in ESL contexts to improve on the delivery system to ensure optimal comprehension.

All in all, the need to conduct a study to investigate the nature of lexical bundles in academic university lectures is imperative to contribute to the understanding of communication. To this end, the present study aims at throwing light on the notion of lexical bundle use in academic lectures of three broad disciplinary sciences, which are life, social and physical sciences. The study has expectations that the results of this study will lead to important implications for language use in particular for university lectures from different disciplines with regard to content comprehensibility facilitated through specific textual features associated with lexical bundles realized in the context of discourse organization. In addition, learning the way lecturers from different disciplinary backgrounds construct lexical bundles structurally and use them to convey specific disciplinary information and those communicative purposes related to their field of study can be of great value for novice lecturers.

1.7 Definition of key terms

This final section provides the definition of key terms to highlight their meaning based on their use in this study.

Formulaic language is commonly defined as a sequence of words (continuous or discontinuous) that is stored in the brain as a prefabricated chunk and retrieved holistically from memory at the time of use (Schmitt, 2004; Wray, 2000; 2002). It operationally refers to any word sequence, syntactically complete and semantically meaningful, which can be looked up in a reliable dictionary (Qi & Ding, 2011). Formulaic language is simply more than two independent morphemes which are stored in a speaker's lexicon as a single unit and is produced without the generative functions of grammar and syntactic rules. In essence, such patterns of language are much more than just string of words which ties with collocational links because much of the communicative purposes of the language depend on the use of these expressions.

Lexical bundles: Lexical bundles fall under the category of formulaic language. Biber et al. (1999, p. 990) defined lexical bundles as “recurrent expressions, regardless of their idiomaticity, and regardless of their structural status,” kinds of word combinations that co-occur frequently and are basically identified empirically, rather than intuitively. These expressions are not idiomatic in nature and do not usually have complete grammatical structures but serve as building blocks of the

language (Biber et al., 1999; Biber & Conrad, 1999). Lexical bundles have been studied under different labels including, formulaic expressions, clusters, chunks, prefabricated patterns, formulas and routines, all referring to the type of multi-word expressions which co-occur frequently and with accidental sequences of three or more words in a given register (*e.g. in the case of the, do you want me to*).

Academic lecture refers to the principal genre of instruction and is a crucial way used to communicate to students the contents of the subject matter as well as other course-related issues (Flowerdew, 1994; Flowerdew & Miller, 1997; Thompson, 1994). It is considered as one of the most important teaching methods in higher education.

Discipline: A discipline is the knowledge of specific subject area shared by the members of the same discourse community (Becher, 1994). Disciplines are categorized according to specific knowledge, for example hard-soft and pure-applied.

Social science: This is defined as the scientific study of human society and social relationships. It focuses on people and how they operate.

Physical science: This branch of science is concerned with the study of inanimate natural objects, including physics, chemistry, astronomy, and related subjects. It focuses mainly on things and how they operate.

Life and medical science: Life science comprises the fields of science that involve the scientific study of living organisms such as microorganisms, plants, animals, and human beings as well as related considerations like bioethics. Medical science is the science of dealing with the maintenance of health and the prevention and treatment of disease.

REFERENCES

- Adel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81–92.
- Aisenstadt, E. (1981). Restricted collocations in English lexicology and lexicography. *ITL. Review of Applied Linguistics*, (53), 53-61.
- Allerton, D. J. (1984). Three (or four) levels of word cooccurrence restriction. *Lingua*, 63(1), 17-40.
- Altenberg, B. (1993). Recurrent word combinations in spoken English. In J. D_Arcy (Ed.), *Proceedings of the fifth Nordic Association for English Studies Conference* (pp. 17–27). Reykjavik: University of Iceland.
- Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word combinations. In A. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 101–122). Oxford: Oxford University Press.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Banerjee, S., & Pedersen, T. (2003). The design, implementation, and use of the ngram statistics package. In *Computational Linguistics and Intelligent Text Processing* (pp. 370-381). Springer Berlin Heidelberg.
- Barber, C. L. (1962). Some measurable characteristics of modern scientific prose. In F. Behre (Ed.), *Contributions to English syntax and philology* (pp. 21-43). Gothenburg: Almqvist & Wiksell.
- Barlow, M. (2000). *Concordancing with MonoConc Pro 2.0*. Athelstan.
- Barr, P. (1990). The role of discourse intonation in lecture comprehension. In M. Hewings (Ed.), *Papers in discourse intonation* (Vol. 16, pp. 5–21). Birmingham, UK: English Language Research, University of Birmingham, Discourse analysis monograph.
- Becher, T. (1989). *Academic tribes and territories*. Milton Keynes: SRHE/Open University Press.
- Becher, T. (1994). The significance of disciplinary differences. *Studies in Higher Education*, 19 (2), 151-61.
- Bex, T. (1996). *Variety in English-texts in society: Societies in Text*. London: Routledge.

- Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. London: Longman.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D. (1996). Investigating language use through corpus-based analyses of association patterns. *International Journal of Corpus Linguistics*, 1, 171-197.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: Benjamin.
- Biber, D. (2009). A corpus-driven approach to formulaic language in English. *International Journal of Corpus Linguistics*, 14(3), 275-311.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286.
- Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora* (pp. 181-190). Amsterdam: Rodopi.
- Biber, D., Conrad, S., & Cortes, V. (2003). Lexical bundles in speech and writing: An initial taxonomy. In A. Wilson, P. Rayson, & t. McEnery (Eds.), *Corpus linguistics by the lune: A festschrift for Geoffrey Leech* (pp. 71-92). Frankfurt/Main: Peter Lang.
- 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., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.
- Biglan, A. (1973). The characteristics of subject matter in different scientific areas. *Journal of Applied Psychology*, 57(3), 195-203.
- Bolinger, D. (1975). *Aspects of language* (2nd ed.). New York: Harcourt Brace Jovanovich.
- Bolinger, D. (1976). Meaning and Memory. *Forum Linguisticum*, 1, 1-14.
- Borg, E. (2003). Discourse community. *ELT journal*, 57(4), 398-400.
- Butler, C. S. (1997). Repeated word combinations in spoken and written text: some implications for functional grammar. In C. S. Butler, J. H. Connolly, R. A. Gatward, & R. M. Vismans (Eds.), *A fund of ideas: Recent developments in functional grammar* (pp. 60-77). Amsterdam: IFOTT, University of Amsterdam.

- Canale, M. (1983). From Communicative Competence to Communicative Language Pedagogy. In Richards, J.C. and Schmidt, R. Eds. *Language and Communication*. London: Longman.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* (1): 1–47.
- Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, 6(2), 5-35.
- Chafe, W. L. (1968). Idiomaticity as an anomaly in the Chomskyan paradigm. *Foundations of language*, 4(2), 109-127.
- Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology*, 14(2), 30–49.
- Cheng, S. W. (2012). “That’s it for today”: Academic lecture closings and the impact of class size. *English for Specific Purposes*, 31(4), 234-248.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: M.I.T. Press.
- Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1), 72-89.
- Connor, U. (1996). *Contrastive Rhetoric. Cross-Cultural Aspects of Second-Language Writing*. Cambridge: Cambridge University Press.
- Conrad, S. M. (1996). Investigating academic texts with corpus-based techniques: An example from biology. *Linguistics and Education*, 8, 299–326.
- Cortes, V. (2002). *Lexical bundles in published and student academic writing in history and biology*. Unpublished Doctoral Dissertation, Northern Arizona University.
- 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., & Csomay, E. (2007). Positioning lexical bundles in university lectures. In M. C. Campoy & M. J. Luzo’n (Eds.), *Spoken corpora in applied linguistics* (pp. 57–76). Bern: Peter Lang.
- Cortes, V., & Csomay, E. (2009). Lexical bundle distribution in university classroom talk. In S. T. Gries, S. Wulff, & M. Davies (Eds.), *Corpus-linguistic*

applications: Current studies, new directions (pp. 153–168). Amsterdam: Rodopi.

Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, 16, 129–147.

Crystal, D. (2004). *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press.

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., (2004). Preferred sequences of words in NS and NNS Speech. *Belgian Journal of English Language and Literatures (BELL)*, 2, 225–246.

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.

Dudley-Evans, T. (1994). Variations in the discourse patterns favoured by different disciplines and the pedagogical implications. In J. Flowerdew (Ed.), *Academic listening: Research perspectives*, (pp. 146–158). Cambridge: Cambridge University Press.

Dudley-Evans, T., & Johns, T. (1981). A team-teaching approach to lecture comprehension for overseas students. The teaching of listening comprehension. London: The British Council, ELT documents special, pp. 30–46.

Duszak, A. (2005). Between styles and values: an academic community in transition. *Identity, Community, Discourse: English in Intercultural Settings*, 69–94.

Erman, B. (2009). Formulaic language from a learner perspective. *Formulaic language*, 323.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *TEXT-THE HAGUE THEN AMSTERDAM THEN BERLIN-*, 20(1), 29–62.

Ewer, J. R., & Hughes-Davies, G. (1971). Further notes on developing an English programme for students of science and technology (1). *ELT Journal*, 26(1), 65–70.

Fillmore, L. W. (1979). Individual differences in second language. In C. Fillmore, D. Kempler, & W. Wang (Eds.), *Individual differences in language ability and behavior* (pp. 203–228). New York: Academic Press.

Firth, J. R. (1951). *Modes of meaning. Essays and Studies* (The English Association), 118–149.

- Flowerdew, J. (1994). *Research of relevance to L2 lecture comprehension: An overview*. In J. Flowerdew (Ed.), *Academic listening* (pp. 7–29). Cambridge: Cambridge University Press.
- Flowerdew, J., & Miller, L. (1997). The teaching of academic listening comprehension and the question of authenticity. *English for Specific Purposes*, 16(1), 27–46.
- Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing*. Harlow: Longman, (pp. 75–93).
- Grabe, W., & Kaplan, R. (1996). *Theory and practice of writing*. London: Longman.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 145–160). Oxford: Oxford University Press.
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational evaluation and policy analysis*, 11(3), 255–274.
- Hakuta, K. (1974). Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning*, 24(2), 287–297.
- Halliday, M. A. K. (1985). *Spoken and Written Language*. Victoria, Australia: Deakin University Press.
- Halliday, M. A. K. (1988). On the language of physical science. In Ghades, M. *Registers of written English: situational factors and linguistic features* (pp. 161–178). London: Frances Pinter.
- Haswell, R. (1991). *Gaining Ground in College Writing: Tales of Development and Interpretation*. Dallas: Southern Methodist University Press.
- Herbert, A. (1965). *The structure of technical English*. London: Longman.
- Herrington, A. J. (1985). Writing in academic settings: A study of the contexts for writing in two college chemical engineering courses. *Research in the Teaching of English*, 331–361.
- Howarth, P. (1998). Phraseology and second language proficiency. *Applied linguistics*, 19(1), 24–44.
- Hyland, K. (2004). Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing*, 13(2), 133–151.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4–21.

- Hymes, D. H. (1967). Models of the interaction of language and social setting. *Journal of Social Issues*, 23(2), 8-38.
- Hymes, D. H. (1972). On communicative competence. In J. B. Pride and J. Holmes (Eds.), *Sociolinguistics: Selected readings*. Harmondsworth: Penguin.
- Ivanič, R. (1998). *Writing and Identity. The Discoursal Construction of Identity in Academic Writing*. Amsterdam: John Benjamins.
- 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, 253–267.
- Jespersen, O. (1917). *Negation in English and other languages*. Copenhagen: A.F. Høst.
- Jespersen, O. (1924). *The philosophy of grammar*. London: Allen & Unwin.
- Karabacak, E., & Qin, J. (2013). Comparison of lexical bundles used by Turkish, Chinese, and American university students. *Procedia-Social and Behavioral Sciences*, 70, 622-628.
- Khuwaileh, A. A. (1999). The role of chunks, phrases and body language in understanding co-ordinated academic lectures. *System*, 27(2), 249-260.
- Kjellmer, G. (1991). A mint of phrases. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics: Studies in honour of Jan Svartvik* (pp. 111–127). London: Longman.
- Kolb, D.A. (1981). Learning styles and disciplinary differences. In A. W. Chickering (Ed.), *The modern American college: Responding to the new realities of diverse students and a changing society* (pp. 232-255). San Francisco, CA: Jossey-Bass.
- Krathwohl, D. R. (1998). *Methods of educational and social science research: An integrated approach* (2nd ed.). New York: Longman.
- Kuiper, K. (1996). *Smooth talkers: The linguistic performance of auctioneers and sportscasters*. Mahwah, NJ: Erlbaum.
- Lattey, E. (1986). Pragmatic classification of idioms as an aid for the language learner. *IRAL-International Review of Applied Linguistics in Language Teaching*, 24(1-4), 217-234.
- Lee, J. J. (2009). Size matters: an exploratory comparison of small- and large-class university lecture introductions. *English for Specific Purpose*, 28(1), 42-57.
- Lewis, M. (1993). *The lexical approach* (Vol. 1, p. 993). Hove, UK: Language Teaching Publications.

- Manes, J., & Wolfson, N. (1981). The compliment formula. In F. Coulmas (Ed.), *Conversational routine: Explorations in standardized communication situations and prepatterned speech* (pp. 115–132). New York: Mouton Publishers.
- Martin, J. R. (1985). Process and text: two aspects of human semiosis. *Systemic perspectives on discourse*, 1(15), 248-274.
- Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach*. Oxford: Clarendon Press.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford University Press.
- Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics*, 11(3), 283–304.
- Paltridge, B. (2006). *Discourse analysis: An introduction*. Continuum International Publishing Group.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. Richards & R. Schmidt (Eds.), *Language and Communication* (pp. 191-225). London: Longman.
- Peters, A. (1983). *The units of language acquisition*. Cambridge: Cambridge University Press.
- Qi, Y., & Ding, Y. (2011). Use of formulaic sequences in monologues of Chinese EFL learners. *System*, 39(2), 164-174.
- 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.
- Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of experimental psychology: General*, 118(3), 219.
- Reppen, R. (2004). Academic language: an exploration of university classroom and textbook language. In U. Connor & T. A. Upton (Eds.), *Discourse in the professions* (pp. 65–86). Amsterdam/Philadelphia: John Benjamins.
- Reppen, R. (2010). Building a corpus. What are the key considerations? *The Routledge Handbook of Corpus Linguistics*. Londres/Nueva York: Routledge, 31-38.
- Schmitt, N. (2004). *Formulaic sequences: Acquisition, processing and use*. Amsterdam: John Benjamins.

- Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 1–22). Philadelphia: John Benjamins.
- Scott, M. (1997). PC analysis of key words—and key key words. *System*, 25(2), 233–245.
- Scott, M. (2008). WordSmith Tools version 5, Liverpool: Lexical Analysis Software.
- Simpson-Vlach, R., & Ellis, N. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 463–512.
- Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
- Strunkyt, G., & Jurkūnait, E. (2008). *Written academic discourse: Lexical bundles in humanities and natural sciences*. (Unpublished master's thesis). Vilnius University.
- Strzalkowski, T. (1999). *Natural language information retrieval* (introduction pp. i–xx). Dordrecht: Kluwer.
- Swales, J. (1988). Discourse communities, genres and English as an international language. *World Englishes*, 7(2), 211–220.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.
- Swales, J. (2004). *Research genres: Explorations and applications*. Ernst Klett Sprachen.
- Tannen, D. (1987). Repetition in conversation as spontaneous formulacity. *Text*, 7(3), 215–243.
- Taylor, S. J., & Bogdan, R. (1998). Working with data: Data analysis in qualitative research. *Introduction to qualitative research methods: A guidebook and resource*, 134–164.
- Thompson, S. (1994). Frameworks and contexts: A genre-based approach to analysing lecture introductions. *English for Specific Purposes*, 13(2), 171–186.
- Thompson, S. E. (2003). Text-structuring metadiscourse, intonation and the signalling of organization in academic lectures. *Journal of English for Academic Purposes*, 2(1), 5–20.
- Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied linguistics*, 16(2), 180–205.

Wood, D. (2001). In search of fluency: What is it and how can we teach it? *Canadian Modern Language Review*, 57, 573-589.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, 21(4): 463- 489.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press.

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: an integrated model. *Language & Communication*, 20(1), 1-28.

Young, L. (1990). *Language as behaviour, language as code: A study of academic English*. Amsterdam: John Benjamins.

Young, L. (1994). University lectures – macro-structure and micro-features. In J. Flowerdew (Ed.), *Academic listening* (pp. 159–176). Cambridge: Cambridge University Press.