FEATURES OF LEXICAL VERBS IN THE DISCUSSION SECTION OF MASTERS' DISSERTATIONS

FAIZAH BINTI MOHAMAD NUSRI

FACULTY OF LANGUAGES AND LINGUISTICS
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

University of Malaya

FEATURES OF LEXICAL VERBS IN THE DISCUSSION SECTION OF MASTERS' DISSERTATIONS

FAIZAH BINTI MOHAMAD NUSRI

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

FACULTY OF LANGUAGE AND LINGUISTICS
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

University of Malaya

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Faizah binti Mohamad Nusri

Matric No: TGB 120045

Name of Degree: Master of English as Second Language

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

Features of Lexical Verbs in the Discussion Section of Masters' Dissertations

Field of Study: Corpus Linguistics

I do solemnly and sincerely declare that:

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

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

University of Malaya

ABSTRACT

This study examines the use of lexical verbs in Discussion sections of Master's Degree Dissertations submitted to the Faculty of Languages and Linguistics of University of Malaya. Lexical verbs are known to be prevalent in academic discourse notably in modulating writers' ideas and stances to the readers. In order to convey their thoughts and communicate well with the readers, precise and succinct use of lexical verbs is paramount. Building on this background, the primary focus of the paper is an investigation of the use of lexical verbs in advanced ESL writing. Lexical verbs found in the academic texts are examined to identify the nature of verbal use and knowledge among advanced ESL speakers by the basis of types of lexical verbs, verb forms and associated verb patterns. It is interesting to note that Mental-Emotive verbs are second most commonly used verbs, after Activity verbs. The results also indicate the writers' tendency to employ passive voice in their writing, which is generally avoided by ESL and non-native writers in other studies despite its conventionalized use in academic discourse. The key verb-based patterns examined in this paper are Passive+PP and V-that patterns in verb-based bundles. Passive voice is predominantly used in combination with PP in *and* by. *This result indicates writers' adherence to academic text objectivity and detachment conventions. Common V-that patterns examplify the writers' confidence in their claims. Analysis of V-that pattern also show writers' tendency to discriminate verb type according to animacy of the subject. This research highlights the use of accurate verb and importance of phraseological knowledge in academic writing and proposes analysis of academic verbs and common patterns as linguistic awareness exercise for advanced ESL learners.*

Keywords: corpus linguistics, phraseology, ESL academic writing

University of Malaya

ABSTRAK

Kajian ini menganalisa kata kerja leksikal di dalam bahagian perbincangan disertasi Ijazah Sarjana Fakulti Bahasa dan Linguistik Universiti Malaya. Kata Kerja leksikal adalah sangat penting dalam penulisan wacana akademik kerana ia dapat membentuk idea dan pendirian penulis supaya dapat difahami sepenuhnya oleh pembaca. Untuk penyampaian idea dan komunikasi yang baik antara penulis dan pembaca, penggunaan kata kerja yang tepat patut diberi fokus. Atas sebab-sebab ini, fokus utama kajian ini adalah untuk mengkaji ciri-ciri penggunaan kata kerja leksikal oleh penulis ESL tahap tinggi. Kata kerja leksikal yang dikaji dalam teks akademik adalah dari segi jenis kata kerja, bentuk kata kerja dan pola kata kerja yang selalu digunakan. Kajian mendapati bahawa pengunaan kata kerja Mental-Emotif adalah yang kedua paling kerap selepas kata kerja Aktiviti. Dapatan juga mendapari penulis-penulis di dalam korpus ini juga lebih cenderung menggunakan ayat pasif dalam penulisan mereka walaupun telah diketahui berdasarkan kajian lalu yang teknik penulisan ini selalu dielak oleh penulis bukan natif. Pola penggunaan kata kerja leksikal pula dikaji dalam bentuk *Passive + Preposition* dan *V-that*. Ayat Pasif paling banyak digunakan dengan kombinasi preposisi *in* dan *by*. Ini menunjukkan pematuhan penulis-penulis kepada konvensi wacana akademik iaitu objektiviti. Pola penulisan *V-that* menunjukkan tahap keyakinan mereka terhadap kenyataan yang mereka. Analisis pola V-that juga menunjukkan penulis mendiskriminasi jenis kata kerja yang digunakan mengikut animasi objek, samaada hidup atau pun tidak. Kajian ini turut menekankan kepentingan penggunaan kata kerja yang tepat dan rangkai kata dalam penulisan akademik dan mencadangkan analisis kata kerja dan pola biasa dalam penulisan akademik sebagai latihan kesedaran linguistik untuk pelajar ESL.

Kata kunci: linguistik corpus, fraseologi, ESL penulisan akademik

University of Malaya

ACKNOWLEDGEMENTS

I would like to thank Almighty Allah for giving me the opportunity, determination and strength to complete my research. His continuous grace and mercy were with me throughout my life and ever more during the tenure of my research. My enormous thanks also goes to my mother who has never stopped praying for my success. I also am deeply grateful for the exceptional help and guidance given by my supervisor Dr. Siti Zaidah Zainuddin without whom this dissertation would not be complete. To my friends who have also provided with strength to persevere through this journey, I could not thank you enough.

University of Malaya

TABLE OF CONTENTS

Abstract iii

Abstrak iv

Acknowledgements v

Table of Contents vi

List of Figures ix

List of Tables x

List of Appendices xi

CHAPTER 1: INTRODUCTION 1

1.1 Statement of the problem 2

1.2 Background of ESL Status in Malaysia 4

1.3 Research questions 7

1.4 Research objective 8

1.5 Scope of the study 8

1.6 Limitation of the study 10

1.7 Significance of the study 10

1.8 Definition of terms 11

1.9 Summary 14

CHAPTER 2: LITERATURE REVIEW 15

2.1 Introduction 15

2.2 Academic discourse 17

2.3 Lexical Verbs 23

2.3.1 Phrasal verb 25

2.4 Study of corpus 26

University of Malaya

2.5 Lexical verbs in academic discourse 27
2.6 ESL Writers and Use of Lexical Verbs 29
2.7 Verb Lemmas vs. Verb Forms 33
2.8 Language Formulaicity 39
2.9 Language Formulaicity; Phraseology and Collocation 43
2.10 Summary 47

CHAPTER 3: METHODOLOGY 48
3.1 Introduction 48
3.2 Analysis of authentic language data 48
3.3 Corpus Linguistics 50
3.4 The corpus 52
3.5 Data Collection Procedures 53
3.6 Ethics of the study 53
3.7 Data Analysis Procedures 54
3.7.1 UCREL CLAWS Tagset 55
3.7.2 WordSmith Tool 6.0 56
3.8 Summary 59

CHAPTER 4: DATA ANALYSIS AND RESULTS 60
4.1 Types of lexical verbs 60
4.2 Lexical Verbs and Verb Forms 68
4.3 Verb forms 70
4.3.1 Tense 71
4.3.2 Aspect 73
4.3.3 Infinitives and gerunds 75
4.3.4 Passive and active voice 76

University of Malaya

4.4 Associated patterns 78
4.4.1 Passive verb + PP fragment 79
4.4.1.1 Short passive and long passive 82
4.4.2 (VP) + that clause fragment 85
4.4.2.1 Tense of Reporting 88
4.4.2.2 Animate vs Inanimate subjects 94
4.5 Summary 95

CHAPTER 5: DISCUSSION, IMPLICATIONS AND CONCLUSION 97

5.1 Introduction 97
5.2 Types of verbs 97
5.3 Usage of verb forms 100
5.4 Analysis of Verb forms 102
5.5 Tense and Aspect 102
5.6 Common verb patterns 104
5.7 Summary of discussion 107
5.8 Pedagogical Implications 108
5.8.1 Academic Wordlist 108
5.8.2 Teaching of Grammar 110
5.8.3 Explicit teaching of Phraseology and Collocation 111
5.8.4 Consciousness-raising activity via data-driven learning (DDL) 113
5.9 Conclusion 116
References 118
Appendix a 129

University of Malaya

LIST OF FIGURES

Figure 2.1 Recurring discourse of academic writing 19

Figure 3.1 Example of text with Discussion subheading 54

Figure 3.2 Example of text with POS-tags 55

Figure 3.3 WordSmith Tool Wordlist 57

Figure 3.4 Example of concordance lines 58

Figure 4.1 Percentage occurrence of lexical verbs in specific verb groups 60

University of Malaya

LIST OF TABLES

Table 1.1 Verb-based structural classification 9

Table 2.1 Parameters of phraseology in corpus linguistics 46

Table 3.1 Summary of corpus 52

Table 3.2 CLAWS Horizontal Output Abbreviation 55

Table 3.3 Stages of corpus analysis 59

Table 4.1 Verb phrase-based bundles and semantic meaning 62

Table 4.2 Examples of Report Verbs with *that* Complementiser 65

Table 4.3 Examples of Logical Semantic Verbs 66

Table 4.4 Concordance Lines with Linking verbs 68

Table 4.5 CLAWS Horizontal Output Abbreviation 69

Table 4.6 Lemmas and Verb Form Breakdown 70

Table 4.7 Verbs in Infinitives and Gerunds 75

Table 4.8 Verb-based Lexical Bundles in Raw Frequency 79

Table 4.9 Passive Verb + PP Fragment in Raw Frequency 79

Table 4.10 Common Report Verbs 85

Table 4.11 Categorization of Report Verbs as Epistemic Verbs 87

Table 4.12 Subject of V_that construction by frequency 94

Table 5.1 High Frequency verbs and synonyms 101

Table 5.2 Stock Phrases for Report Verbs 106

University of Malaya

LIST OF APPENDICES

Appendix A: Wordlist of the corpus sorted by frequency……... 129

University of Malaya

CHAPTER 1: INTRODUCTION

Thesis writing is considered one of the most formidable tasks for many graduate students; even more so when the task needs to be written in a second language. Despite the expanding number of English as a Second Language (ESL) writers, many still consider the task as the ultimate challenge. ESL writers have commented that they could not seem to achieve the same rhetoric and linguistic repertoire as read in journal articles and academic writing, most written or edited by native speakers (Karim Sandengi, 2015). The writers are not the only ones suffering in this regard, however. Similar sentiments have been echoed by both English for Academic Purpose (EAP) and English instructors as they struggled over the cryptic and ambiguous constructs produced by non-native writers (ESL or EFL (English as Foreign Language).

There are several contributing factors that could have caused this linguistic conundrum and one of them is the choice of lexical verbs. A study by Bloch (2010) has identified that non-native writers have had difficulties in choosing the verbs that could "satisfy both syntactic requirements of their sentences and, perhaps more importantly, to express their attitudes towards the claims" (p. 221). In many ways, choice of lexical reflects the writers' stance and attitudes in their arguments. Limited exposure and understanding of the subtleties of verbs uses may have shaped this condition into what it is now. Observations of grammatical choices in rhetorical context have allowed writers to examine and pinpoint certain grammatical aspects that cannot be learned, but rather observed as a system of abstract rules. Realizing this, an examination of a large body of text (the corpus) has become a powerful pedagogical tools, especially for advanced learners. This study intends to describe features of lexical verbs by identifying categories of verbs used in the ESL corpus created specifically to serve above-mentioned purposes. The examination of verb forms could also provide a deeper analysis of tenses and aspects, voice and

University of Malaya

associated patterns favoured by the writers and common associated patterns of lexical verbs.

1.1 Statement of the problem

Over the course of ESL teaching and learning, learners will come to learn that the verb is perhaps one of the most crucial elements in any forms of text constructions (Quirk et al., 1985, as cited in Hinkel, 2004). Lexical verbs, especially, are embodiments of expressions of action, state and meaningful predicates in texts (Crystal, 2003). Granger and Paquot (2008, p. 1) argue that lexical verbs play predominant parts in EAP functions such as "expressing personal stance, reviewing the literature, quoting, expressing cause and effects, summarizing and contrasting". The use of proper verbs also allows writer to argue their stands and establish their position within the related literatures. Argument structure, in any genres, is an abstract thought illustrated by varying grammatical constructs that must include a verb (Goldberg, 2013). The choice of verbs, in particular in non-native writing has been noted to be the main point of distinction from their native counterpart (Burrough-Boenisch, 2003). Verbs have been classified into quite a number of categories by researchers, but among the various categorization of verbs, 'reporting verb' has been under the most scrutiny (Charles, 2006; Zhang, 2008; Bloch, 2010). In general, EAP studies have tended to focus on one specific category of verbs rather than give a general overview of the use of lexical verbs in academic discourse. These studies, though undoubtedly enlightening, have yet to paint a complete picture of how learners employ lexical verbs in academic writing.

The present study reviews the use of verb categories and verb forms in Discussion sections as they are where the writers' voice is the strongest. This is because in this specific section, writers are required to make statements of results, link results to previous

University of Malaya

research, explain findings and make deductions and hypotheses among other things (Swales, 1990). These processes needless to say depend on the writers' ability to provide a personal yet convincing arguments that support their findings. Previous studies have remarked the challenging nature of the constructing the Discussion sections (Belcher, 2009; Parkinson, 2011 as cited in Safnil, 2013). Belcher (2009, as cited in Safnil, 2013) comments that Discussion section is the most difficult yet important part of an academic paper that proper arguments need to be structured with care or it will lead to rejection or misinterpretation on the part of the reader.

Despite ample studies of corpus-based and corpus-driven, the study focusing on Discussion sections is quite scarce. To the author's knowledge, there has yet a published paper focusing on Discussion sections with the exceptions of a number of writing guide and how-tos. In Discussion sections, the writers are required to report, cite, argue and defend their findings where verbs play a dominant, but complex role; whether in lexical choices (*say* or *state*), verb forms (*mention* or *mentioned*) and structures (*Hunston argues* that *or It has been argued*). Such nature of verbs can be problematic to ESL writers as academic constructs in particular tend to employ highly conventionalized tense and aspects (Swales, 1990). For example, Swales finds that while the perfect aspect can be employed in certain chapters of a paper, such as in introduction and literature review chapters, the progressive aspect is rarely, if ever, used in an academic prose. For these reasons, Min (2013) adds that the usage of English verb tense and aspect is notorious for being confusing and difficult for non-native students to learn and for teachers to explain, not only because it is complicated in its nature, but also because linguists and grammarians have defined, categorized, and described them in such diverse ways.

Corpus-based research focusing on examining and exploring ESL leaners' lexical knowledge is still very small in number considering the long-standing interest by the

University of Malaya

academic community. Despite the increased attention in the past two decades and continued interest in the matter on the global scale (Nesselhauf 2003,2005; Granger and Paquot 2008; Henriksen, 2012), there seems to be very little response from within the Malaysian linguistic community. A thorough search corpus–based for and corpus-driven studies has yielded only a handful of results consisting of Kamariah Yunus and Su'ad Awab's (2011) 'Investigation of Collocational Competence among Law Undergraduates' at a local university and Hong et al. (2011), a corpus-based 'Error Analysis on Malaysian Students' Writing'. In addition to those, a study of reporting verbs in M.A theses was conducted in a local university by Manan and Mohd Noor (2014) who found that even advanced ESL writers tend to repeat the use of certain type of verbs. At this point, a majority of the research have chosen to examine collocational errors by novice to intermediate learners of English and only a very limited number has attempted to examine language use of advanced learners with the exception of Manan and Mohd Noor (2014).

As an effort to tackle these problematic areas, a study on phraseological patterns of verbs could be a starting point. The learning of highly recurrent verb patterns that are specific to academic prose can be seen as a means to a productive approach to teaching and learning salient verbs. This study intends to analyse advanced learners' corpus that encompasses verbs used in academic writing that include the analysis of verb types, forms and associated patterns.

1.2 Background of ESL Status in Malaysia

The English Language status in Malaysia has come a long way since before our independence. During the British occupation, the English language was utilized as a medium of instruction in a majority of the public schools. As Malaysia gained her independence, the new government under the New Education Policy has taken a brave

University of Malaya

step to convert the English schools into national schools. The conversion to national schools took a gradual pace. It began in 1970 and ended in 1983, with the repeal of the Lower Certificate of Education (LCE) in 1978 and the Malaysian certificate of Education in 1980. It is a crucial step to promote national unity in a multilingual society through the use of Bahasa Malaysia. At the same time, English language is demoted to a second language status due to its status as the world's lingua franca, which, if mastered, would allow the nation to be globally competent especially in the economic sector. This is because to operate competently in an open world market, competent use of communicative English is necessary. Accordingly, English language functions as a strong second language; "not a native language of the country but used for certain purposes and by certain people within the country" (Citravelu, Sithamparam and Teh, 2005, p. 12).

The ESL status in Malaysia has been a subject of much discussion among many scholars, as a result of the ever changing educational policy. In line with this, the Malaysian Education system has allotted 300 minutes of English period in a week (under the KSSR programme) only second to Bahasa Malaysia with 360 minutes in a week which shows, if nothing else, that English is considered important as the nation's second language. The aims of the English Language Curriculum for Primary schools are as follow;

"The English Language Curriculum for Primary Schools aims to equip pupils with basic language skills to enable them to communicate effectively in a variety of contexts that's appropriate to the pupils' level of development".

(Curriculum Development Division, 2011).

University of Malaya

One of the many scenarios that essentially explains the constant challenge to ESL status in Malaysia is described in the following anecdote. It would seem like ideally, anyone who has gone through six years of primary education, five years of secondary education and at least one semester of mandatory English language class in tertiary education should have guaranteed English competency among Malaysians. On the other hand, it seems like English has never been the nation's second language, or maybe even a third, due to a number of handful other languages that are spoken and taught within their home and school environment. In Chinese and Tamil secular schools for example, English period is reduced to 180 minutes weekly. It would be much more interesting to observe learners in the rural areas whom exposure to the language is only limited to the classroom environment and possess very limited language skills. Adding to that is the continuous concern over university students' English competency which does not at all reflect the time and effort put into helping them to master the language.

Despite many arguments regarding the decline of English competency and proficiency in Malaysia, Malaysians at large show higher proficiency compared to a foreign speaker of the same language. According to the English Profiency Index (2017) website, Malaysia ranks second in Asia and thirteenth among 80 countries. Furthermore, English has been accorded to cover more communicative domains in Malaysia, especially in terms of being a medium of education and advertisements. Nowadays, English is also seen as a crucial tool for higher education and employment and thus a key to rise up the socioeconomic ladder. Based on these observations, we can say that English is still a second language in Malaysia.

University of Malaya

1.3 Research questions

The research questions are as follows;

1. What are the types of lexical verbs used in Discussion sections in M.A dissertations?
2. To what extent are the verb forms used in Discussion sections in M.A dissertations?
3. What are the common associated patterns of lexical verbs used in Discussion sections in M.A dissertations?

For the first research question, Top 100 verbs in term of frequency are categorized following Hinkel's classification of verbs in academic writing. They are activity verbs, reporting verbs, mental/emotive verb, linking verbs and logico-semantic relationship verbs. This is achieved by using Wordsmith Tool software to list all the tokens by frequency and manually identifying the Top 100 verbs. Therefore, each verb will account to 1% of the overall percentage. The second research question requires analysis of the verb forms. To answer this, CLAW Tagset 5.0 is used to label all part-of-speech (noun, verbs, adjectives etc). The verbs (tagged VV*) by the software are later categorized into further classification (VVN, VVG and so on). Analysis by the basis of frequency will show which verb forms are most favoured by the population of the corpus. Examination of verb forms will also show usage of tense and aspect through analysis of verb forms labelled VVB (base form), VVD (past tense), VVZ (present tense) and VVN (past participle). The third research question is answered following the Verb-based structural classifications by Biber et al., 1999 (see Table 1.1) which has been used to examine word patterns. Frequencies of these patterns are recorded and select patterns are examined further.

University of Malaya

1.4 Research objective

The main objective of this study is to find out how lexical verbs are used in discussion sections in M.A dissertations. The verbs are examined in three aspects; types of verbs, usage of verb forms and patterns in which the verbs are most commonly found. It is not the study's aim to pinpoint errors made in verbal usage since the corpus' population is of advanced L2 users. Rather the aim of the study is to highlight the features of verbal use by Malaysian ESL learners, with the benefit of empirical data and discover distinctive aspects of use that set apart our local constructs from the natives. The findings from this research can also benefit the growing numbers of ESL academic writers to be more proficient and articulate in the academic discussion.

1.5 Scope of the study

Verbs have been classified in many ways notably by Biber et al. (1999) and Hinkel (2004). This study follows Hinkel's classification of verbs namely activity verbs, reporting verbs, mental/emotive verb, linking verbs and logico-semantic relationship verbs. Verb forms are discussed in relation to tense, aspect and voice. These covers the present and past tense, including the present perfect and past perfect. Passive and active voice are a major point of analysis as the voices have somewhat contradictory opinions regarding choice of voice in academic genre.

This study also identifies the most recurring patterns involving lexical verbs from the corpus in order to answer RQ 3. This area of analysis, recurring patterns, is often linked to the study of phraseology. This is because the patterns involving lexical verbs are more often more exclusive in nature, in the sense that they appear more often in strict patterns with little rooms for deviations. The patterns are broadly referred to phraseology. The patterns that emerge from the data of this study could be an interesting avenue for

University of Malaya

research. Phraseology is a broad subject to explore and the different terms/approaches/ methods may yield varying results. To limit the scope of the study, not all types of phraseology are explored. The term phraseology itself is rather difficult to pinpoint but this paper will attempt to define it in accordance to the focus of this paper. The term is further discussed in Literature Review chapter. The phraseological patterns found in the corpus are classified based on the structural classification of lexical bundles, with focus on verb-based structures introduced in the Longman Grammar of Spoken and Written English (Biber et al., 1999). The structural classification has been widely used in previous studies on word combinations and patterns (Cortes, 2002, 2004; Charles, 2006, Hyland, 2008a, 2008b, Chen and Baker, 2010). Table 1.1 shows the verb-based structural classification (Biber et al., 1999).

Table 1.1 Verb-based structural classification

1	copula *be* + NP/AdjectiveP
2	VP with active verb
3	anticipatory *it* + VP/adjectiveP + (complement-clause)
4	passive verb + PP fragment
5	(VP +) *that*-clause fragment
6	(verb/adjective +) *to*-clause fragment

Note that the structural classification has been modified to cater to the objectives of this study. Therefore, Verb-based clause with copula *be* + NP/AdjectiveP , anticipatory *it* + VP/adjectiveP + (complement-clause) and (verb/adjective +) *to*-clause fragment are not examined since they are mainly formed with *be* or copula verb. The use of active verbs, which also denotes tense markers provides a wealth of research opportunities on its own. However, to fulfil the aim of this study, it will proceed to explore certain patterns of lexical verbs. RQ 3 focuses on two patterns/ bundles which are VP+ that and VP+ PP (Prepositional Phrase).

University of Malaya

1.6 Limitation of the study

The study examines the use of verbs in academic writing, specifically dissertations submitted to the Faculty of Languages and Linguistics which have been awarded the Degree of Master of English as Second Language. The focus of examination is limited to Discussion sections that are included in Chapter 5: Conclusion. This study views its limitations in two major points; corpus and methodology.

As mentioned previously, the selection of participants for this study is very particular and limited. Only 35 dissertations are selected and the writers share common ESL background. Therefore, any future analysis should be done with this particular limitation in mind. In terms of methodology, corpus linguistics, while serving its purpose suitably for the present study, is also arguably riddled with shortcomings. However, considering that its purpose is to record occurrence of collocations, the only concern that should be clarified is the one voiced by Howarth (1996, p. 93). He views that this method is indeed hindered by technology, as it is 'over-dependent on computational method'. Consequently, it loses certain values that come with real-world qualities. However, these shortcomings should not undermine the value of corpus-based linguistics. Rather, they should serve as points of consideration and accountable for researchers and studies. That is to say that researchers to provide an analysis that is comprehensive that also takes into account other factors that could have affected the data rather than focusing solely on the data itself

1.7 Significance of the study

The advantage of corpus-based study such as this has enabled large quantity of natural language data to be processed in efficient and accurate method. This is especially true whereby large quantitative data provide generous insight into linguistic aspects in

University of Malaya

any manner of linguistic research. Also parallel with advancement of computerised technology, corpus-based methodology has created a new pathway to linguistic research, with underlying theoretical approach to the subject (Leech, 1992).

For this study, the corpus collected is Discussion sections in M.A dissertations which are written by ESL authors. By using language processing software such as the CLAW Tagger and WordSmith Tool, the study is able to examine specific word class (verb) and their usage in both forms and patterns. The tokens amount to over 100, 000 units will explicitly provide an evident picture of how these discussion sections are shaped in relation to lexical verb usage. Previous research in this area have been more interested in looking at a specific type of verb, the report verbs (Charles, 2006; Zhang, 2008; Bloch, 2010). The significance of this study is that it should be able to present a comprehensive analysis of how verbs are employed as tools for discussion. The findings from this study can be utilised to bring to attention the vast learning opportunities of data-driven learning and to promote proficient verb usage in academic writing.

1.8 Definition of terms

In this section, this paper will attempt to clarify some of the terms used. Some of terms for example lexical verbs and phraseology appears to be too broad a notion to be tackled in such a simple research. Nonetheless, it is prudent that these terms are clarified to avoid confusion to the readers. Two of these terms are lexical verbs and phraseology.

Generally, lexical verbs can be described as full verbs or open class verbs excluding auxiliary verbs which enable writers to express action, state emotion and predicate meaning in a sentence of written text and there are literally dozens of common lexical verbs in English language (Biber *et al*., 1999), for instance 400 different verbs can

University of Malaya

be generated from 20 times per million words which include daily terms such as *fall*, *choose*, *throw* and *pull*. Many studies on lexical verbs have been published in language corpora decades ago and most corpus-based linguistic studies (Altenberg, 1998) have been carried out during those times. Although corpora are important for providing natural paradigms of grammatical features or words in context, corpus linguistics provides an identical perspective due to the use of quantitative analysis that enables researchers to observe language patterns impossible to be ascertain before (Biber and Conrad, 2001).

Further examination of the lexical verbs leads the analysis of the verb forms. This is especially important if one is interested in having a thorough look into the use of verbs. to The English language has little in variety in term of verb forms, although they pose quite intimidating element to navigate for non-native users. There is the base form (write), third person present (writes), past tense form (wrote), infinitive form or the –ing form (writing) and the past participle form (written).

It has been established that certain words tend to co-occur with a specific type of word which when observed, seem to create patterns of language use. This study intends to scrutinize patterns that are commonly associated with verbs, as outlined by Biber et, al. (1999). These are patterns that are highly salient in academic writing, for example the combination of (VP +) *that*-clause fragment passive verb + PP fragment. Therefore, there are two patterns explored in this paper.

Another terms that is difficult to pinpoint is phraseology. This term has been under scrutiny and criticism by many researchers for as long as the matter has been under study (Gries, 2008). Gries (2008) further argues that lack of effort in clarify these terms is a cause for concern. The effort for clarification of these terms may influence researchers and in future research in such a profound way.

University of Malaya

The term phraseology can be used in distinctive ways by various scholars but various terminologies can be used to interpret the same notion or similar word of co-occurrence for instance the terms *recurrent word combinations* (Altenberg, 1998), *clusters* (Hyland, 2008), *n-grams* (Stubbs, 2007), *phrasicon* (De Cock *et al*., 1998) and *lexical bundles* (Biber and Barbieri, 2007). These terminologies (*recurrent word combinations, clusters, n-grams, phrasicon* and *lexical bundles*) actually predicates to continuous sequences of words acquired through corpus-driven method with distribution criteria and specified frequency and the sequences are utilized by the native language within specified contexts and are fixed multi-word units that possessed customary discourse functions or pragmatic functions (Granger and Paquot, 2008; Chen and Baker, 2010).

Going back to the issue at hand, the previous examples have shown how the terms are used almost interchangeably. As elaborated previously, there are several terms that researchers have used to describe these 'chunks of words'. As mentioned earlier, a number of previous studies have used the clusters, recurrent word, phrasicon, n-grams, lexical bundles and formulaic sequence to refer to phraseology and one of the most prominent term is *collocation*. However, for the purpose of this study, the term phraseology will be used as the umbrella term to refer to these lexical combinations. Phraseologism is defined by Gries (2008, p.6) as the 'co-occurrence of a form or a lemma of a lexical item and one or more additional linguistic elements of various kinds which functions as one semantic unit in a clause or sentence and whose frequency of co-occurrence is larger than expected on the basis of chance'. He also outlines six elements that need to be taken into consideration when classifying phraseology under the paradigm of Corpus Linguistics which will be discussed in Chapter 2.

University of Malaya

1.9 Summary

This chapter has outlined the statement of the problem, aim of the study, research questions, the limitation and the scope of the study. It has also provided a brief summary of ESL status in Malaysia. In the introduction, it is revealed that ESL writers face problems to construct rhetorically and linguistically competent academic prose. Studies have shown that choice of verbs is key to address these problems and the method requires analysis of their constructions in rhetorical context. The focus of this study is contextual uses of lexical verbs, or in other words a corpus-based analysis of lexical verbs of an ESL corpus which allows for further examination of tense and aspect, voice and also associated patterns. This study consists of five chapters: Chapter 1 is the introduction and it provides an overview of the study. Chapter 2 reviews previous studies related to the present study and discusses the definition of certain operational words. Chapter 3 describes the methodology of data collection, procedures taken and how data will be analysed and categorized. Chapter 4 consists of the results and discussion of the results. Chapter 5 includes a summary of the results and implications for future research and teaching.

University of Malaya

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

English language is a globalized lingua franca which has always been a significant motivating factor in learning and particularly in literature writing (Murugesan, 2003); while simultaneously playing fundamental roles in binding a multicultural nation. It is utilized extensively in all aspects of daily lives, from education to business matters. In Malaysia, English is utilized as an additional language to be in the education system in all levels namely primary, secondary and tertiary; as well as for professional global contacts of various sorts (Darmi and Albion, 2013). Malaysians mostly, even though proven to have obtained appropriate English education and excelled in national proficiency examinations; somehow due to differences in attitudes (Gobel *et al.*, 2013), would not apply what they have learned into daily communication and academic writings. It is not surprising to encounter such issues since it is a very common problem for countries who utilize English language as a second or foreign language (Souriyavongsa *et al.*, 2013).

According to studies conducted by Samsiah and Aishah (2001), when inquired regarding their English language proficiency, the respondents (who are of supporting staffs at three public universities in peninsular Malaysia) mostly described their English speaking skills as the weakest, followed by reading and listening skills which they expressed as satisfactory respectively. English language is only to be utilized in daily duties such as records, filing and documentation, public relations, and communication with outsiders; which signifies that they do not really use English unless for writing purposes, administrative and academic records.

Malaysian students, graduates and workers in reality do possess relatively moderate and admirable English language proficiency, and it is well proven when it

University of Malaya

comes to their writing skills. They, without questions, passed their mandatory written English proficiency examinations but would face difficulties when it comes to daily communication and academic writing especially with in-depth use of lexical verbs. Students learning English as a second language have experienced a wide range of contexts in which they have acquired their mother tongue, but have a much more restricted range of contexts in English (Tabors, 1997).

In addition, non-native speakers of English have shown interest and determination to compete in the global market by acquiring English as a lingua franca to ensure their voices were heard by the developed countries while still utilizing and keeping their mother language or home language for unison. As once an official statement by the Chilean government stated that the quality of English we used nowadays may not be challenge the quality of the Shakespeare's, but the important thing is the understanding of English and to utilize it as a tool of communication in the real world especially in academic writing (Rohter, 2004).

Throughout the globalization of economy and academics, we are interacting between cultures in which the importance of learning a second language becomes very significance to ensure strong communication and expand abilities and opportunities in the real world. Scientific studies also prove that learning a second language enhance brain activity and stimulates creativity allowing us to understand different cultures across the world and gain more appreciation of human society and the diversity of culture. Nowadays, a great demand for English as a second language have risen as English become the gateway to the world of knowledge, literature, culture and commerce providing access to the real world in a way other language could not. English has become the key to the involvement in the global conversation and has become world's most widely spoken second language surpassing other languages.

University of Malaya

This section generally discussed the whole viewpoint of the research from the most fundamental issue to the related and relevance points supported by previous study by researchers and published facts. This section is divided into five subsections which consisted of clear description and definition of academic discourse and lexical verbs which is the vital point of this study. This chapter is further described and explained on the utilization of English as a second language in academic writing, listing out and discussed about lexical verbs in general and further focuses on lexical verbs in academic discourse. This section then elaborates on verb lemmas and verb forms, to define how big the role of lexical verbs in English as a second language in academic writing. This is done by reviewing the studies that are related to English as a second language writers and their use of verbs in academic writing. Lastly, all of the issues, facts and ideas stated in the subsections are summarized at the end of the chapter.

2.2 Academic discourse

The interdisciplinary study of academic discourse focuses on the successful communication among academic and non-academic discourse and their speakers and also successful communication between members of the academic institution or communities in distinguish cultures (Bennet, 1991). Academic discourse provides authors with the language tools that involved syntax and vocabulary that is essential to interpret the content using complete sentences which further allow authors to build structured texts and dialogues with meaningful context in appropriate academic language (Bennet, 1991; Barton, 1994; Katnic-Bakarsic, 2004). Academic discourse can be defined as the framework of rhetorical strategies of written academic texts with or without the presence of expressive features and additional style of interpreting the idea and presenting the whole texts.

University of Malaya

In some studies and the viewpoints of critical language stylistics and linguistics, the language utilized in academic discourse is considered as the language of authority and power which allow distinction between academic writing and different cultures of language, allow establishment of successful communication especially between writers and readers (Katnic-Bakarsic, 2004; Hyland, 2007). From the point of view of those interested in discourse studies, social semiotics and critical discourse analysis, the neutrality and objectivity of academic discourse are superficial while traditional linguists argue with the point of view and stated that academic discourse is a neglected stylistic in language studies and the fact that it is completely neutral category of discourse. Several authors constitute distinction in academic discourse created two kinds of readers which are outsiders and insiders and further explore facts in stylistic of academic discourse (Katnic-Bakarsic, 2004).

The terminology of academic discourse could mean a broad range of oral and written genres in which some of the content may be close to administrative discourse, communication discourse and also news discourse and essays writing. It may also consist of conflict, tensions and contradiction such as academic discussions, doctoral dissertations and also colloquia. Meanwhile, scientific writing had been categorized as a type of narrative texts consist of sequence of events with emblematic narrative arrangement interpreting the whole content properly (Barton, 1994).

A study by White and Lowenthal (2011) on academic discourse and the development of an academic identity stated that the academic success of university students is largely influenced by their willingness to learn and apply academic discourse and their exposure towards the functionality of academic discourse which is expected by the academy that their graduates learn appropriate academic literacy. However, individuals that adopted English as a second language may view academic discourse as

University of Malaya

complicated because people tends to cling their ways to their native language although they acknowledge the social pressure and academic pressure in adopting new discourses (Gee, 2005; Gibson, 2005; White and Lowenthal, 2011).

Academic writing is a common activity and noted as integral culture at university however in a study by Hyland (2013), he argued that academic literacy is vital to every activity done in a university and that specialist forms of academic writing are all that matters in an academic institution. He stated that lecturers and students must increase their fluency in the conventions of English academic writing in order to effectively develop their learning, establishment of their careers and also to further understand their disciplines. One of the conventions mentioned by Hyland (2013) is the capability of a writer to apply the socially suitable features of marking stance in their academic discourse or discourse communities which can be further elaborated as the process of developing oneself identities through socially obtainable discourses (Davies and Harre, 1990). Academic discourse is also known for its conventionalised style of writing, which features recurring characteristics. Four moves are highlighted as the recurring discourse of academic writing which are outlined in the Figure 2.1 below;

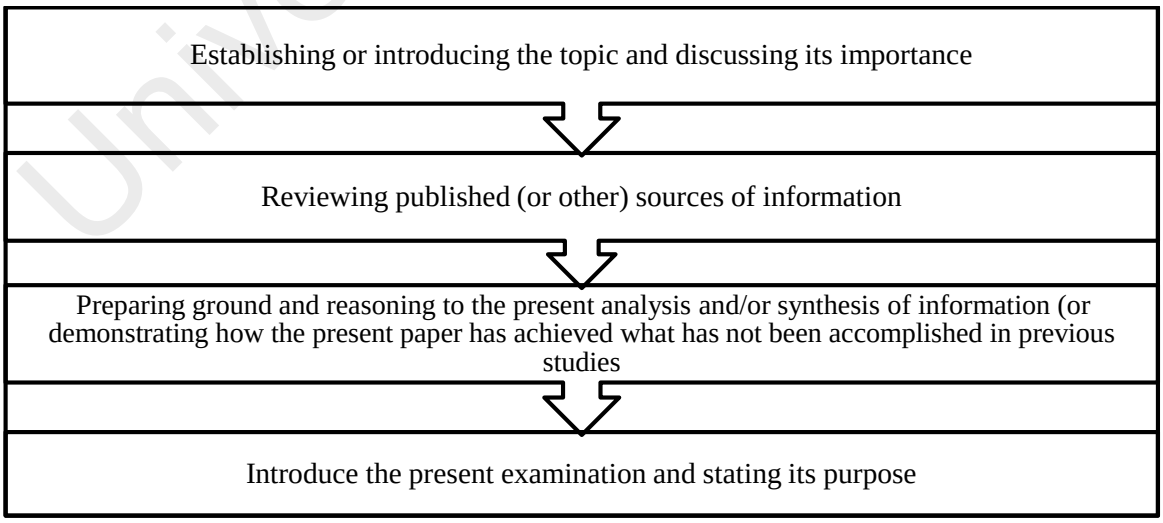


Figure 2.1 Recurring discourse of academic writing

(Source: Johns, as cited in Hinkel, 2004).

University of Malaya

Out of the four moves, research has suggested that writers, especially non-native (NNS), are finding it most difficult to reason and argue their analysis and state their findings in assertive manner. Hyland (1998) finds that non-native writers often face problems in maintaining self-assured position when defending and arguing their stance. It is also noted that non-native writers tend to be vague and avoid making claims that requires 'interpersonal and persuasive effects'.

Academic writing have always been related to postgraduate dissertations writing which sometimes not appropriately organized in terms of interpreting credible representations of themselves in their dissertation and Hyland (2002b) stated that most writers build their identities that are not supported by the discourses communities and out of practices or disciplines which brings to the issue of dissimilarities between writing disciplines and writing practices due to limited exposure to the actual discourse applied by the community.

In consideration of positioning themselves appropriately in conjunction to their work, students shall be disposed to explain the characteristics of stance marking explicitly. This indicates trying to comprehend the practices of real students interacting in real disciplines by portraying and analyzing appropriate texts. Hyland (2013) supports with an idea that states students can only systematize supports, point out collegiality and discuss agreement by making morphological characteristics which combine their texts with their disciplines. It is very important that it is cultivated through the extent of features for instance, writers can locate themselves the nearest to their work by applying the exclusive personal pronoun (I) or position themselves from their work by utilizing express point of view creation or the third person frame of reference. It was stated that the posture a writer believe, portrays the article of faith and epistemology of the discipline they come from (Tang and John, 1999; Stapleton, 2002). This kind of belief developed from the

University of Malaya

viewpoint that written academic communication makes a bombastic appeal to reader, pursuing them to take the writer's viewpoint alternatively than simply stating non-committal facts (Myers, 1990; Tang and John, 1999; Matsuda and Tardy, 2007) however this field of study is still underestimated, in which Hyland (2005) agree to the point that issues regarding writer stance is still new in academic research.

Writers' posture is connected with the concept of averral and attribution. With reference to averral, writers are made-believe to aver the whole propositions in the text and therefore taking the charge for their veracity, except if they are attributed somewhere else (Hunston, 200). Additionally, when an attribution is fashioned, a proposition is held on to a source apart from the writer and amenability is assigned to that individual or being. Therefore, it is the writer who selects whether, to which sources and when to attribute propositions. Allegedly writer stance in academic writing rests as an inadequate understood field. It is not comprehensible how writers shall assimilate their own attitudes, feelings, value appraisal or assessments in the texts that they come through. Nevertheless, the process of writing wraps up in creating a text that we consider the reader will recognize, be aware of and expect, and the process of reading includes drawing on acquisition about what the writer is trying to do (Hyland, 2013; Maroko, 2013).

Each of the academic text is written to be both understandable and accustomed and while either one of the goal is ever completely and absolutely certain, writers who can strongly presume something of what their readers will distinguish of their subject and anticipate of its demonstration are more likely to be persuasive. Hinged on their presumption of their readers and their former experiences with correlative texts, writers continually monitor their expansion discourse to forward these expectations, making eloquent choices which adjudicate proper explicitness and engagement. They identify where greater elaboration or exactitude is required, where readers will require help in

University of Malaya

adapting points, where descriptions or examples are needed and so on. By way of explanation, arguments are predicated and supported by small affectations of propositional elaboration which serve to expand understanding, configure meanings more correctly to the writer's goals, and connect statements to the reader's knowledge-base, experience and processing needs (Hyland, 2007).

The latest research has highlighted that disciplines have different research practices, different views of knowledge and different sight of seeing the world and that these distinctness are portrayed in variety of forms of expression and argument (Hyland, 2000). Approximately, academic writing is not a single associated body of matter but an assortment of subject specified literacies. As a consequence, these literacy members of disciplines associate with their peers and students with their professors. Significant words they choose are ought to present their ideas in a variety of ways that make certain sense to their readers and generally involves adopting an appropriate identity. It is concurred with facts that most of the thing that we write conveys something about us and the variety of connection that we want to compose with our readers. Most unmistakably, anyhow, a writer's identity is constructed by and displayed through the adoption or the absence of the *I* pronoun (Hyland, 2002a).

The means of learning to write at university usually involves the mechanism of creating a new identity (Fan Shen, 1988) which corresponds to the anticipations of the subject teachers who represent a student's new discipline. The author's explicit expressions in a text, or its absence works to generate a reasonable academic identity, and a voice with which to offer an argument. Composing such an identity, nevertheless, is commonly very difficult for second language students. This is partially because these identities may vary considerably from those they are well-known from their everyday lives, or past learning experiences (Cadman, 1997), and also because the reason that

University of Malaya

students are not often taught that disciplinary conventions di.er (Lea and Street, 1999). Concisely, if we utterly assume that academic writing is generally impersonal, we counterfeit variability, and this could have the consequences of restraining our students from approaching terms with the significant needs of their disciplines. Instead of supplying learners with the linguistic ways to achieve their bombastic invisibility, then, we are required to lead them towards knowledge of the alternatives that academic writing offers. The outcomes suggest that academic writing is not the formal faceless prose as it is normally thought to be, but demonstrate considerable differences between disciplines (Hyland, 2002a).

2.3 Lexical Verbs

Generally, lexical verbs can be described as full verbs or the open class of verbs excluding auxiliary verbs which enable writers to express action, state emotion and predicate meaning in a sentence of written text and there are literally dozens of common lexical verbs in English language (Biber *et al*., 1999), for instance 400 different verbs can be generated from 20 times per million words which includes daily terms such as *fall*, *choose*, *throw* and *pull*

Additionally, the qualitative interpretation of corpus analysis allows writers to evaluate the function of lexical verbs for instance, the verbs *get* have various function while the verbs *say*, can only be apply for single function and it is related to an activity. It is perceptible for learners to be exposed to both verbs as they may hear it and use it for daily interaction and communication among members of the community and the native speakers but most grammar books and references for English as a second language learner does not cover these verbs. As an alternative, most books will introduce activity verbs such as *study*, *travel*, *run*, *work*, *play*, *eat* and *like*. Although these activity verbs

University of Malaya

encompass more tangible meanings relating to activity, these verbs are much less familiar. Therefore, even simple quantitative analyses can offer imperative information that material writers and teachers can make use of to revise existing lessons materials (Biber *et al*., 1999; Biber and Conrad, 2001; Swales, 2004).

Furthermore, many researchers had been conducting corpus-based studies in a past few years in order to analyze connections between lexical items and grammatical features. Hence, it is a comparatively straightforward matter to verify whether there are connections between particular verbs and progressive aspect. In reality, especially during communication among members of discourse community, day to day conversation may involved a few lexical verbs for instance *moaning*, *kidding*, *joking*, *starving*, *shopping*, *chasing* and *bleeding* which applies most of the time along with progressive aspect. Nonetheless, the use of norm in daily conversation allowed people to express verbs with the simple aspect. On the other hand, distinctive to the expectations generated by many popular grammars, verb phrases such as *is always telling* and *I've been having* are exclusions rather than the rule (Biber *et al*., 1999).

According to continuous studies by Biber (1988) and Biber and Conrad (2001), they illustrate the implications and insinuation of quantitative analysis towards pedagogical practice and further found out that corpus analyses provide better description and complex elaboration on lexico-grammatical information. They studied related field on the connections among grammatical features that tend to differentiate sets of words for instance nouns controlling to-clauses vs. that-clauses, adjectives, and the most common verbs. Academic discourse further analyzed corpus perspective in terms of the factors that favoring and influencing the preference of as a relative pronoun and preference among grammatical variants. Besides, in order to entertain new types of research questions regarding several issues that had formerly been considered intractable,

University of Malaya

quantitative corpus-based techniques have been suggested by sociolinguists and they had long recognized that linguistic co-occurrence is fundamental to a perceptive of register variation, however they required more research techniques and approach to further identify and clarify the occurring linguistic characteristics. Quantitative corpus research nowadays has filled this breach in knowledge, by utilizing multivariate statistical methods to classify basic of co-occurring linguistic characteristics and to investigate the differences and similarities among registers with respect to those dimensions (Biber and Conrad, 2001; Granger & Paquot, 2015).

2.3.1 Phrasal verb

Previous studies showed that phrasal verbs are notoriously complex for non-native learners to understand in which problems exacerbated and arise in the English language as a whole (Biber *et al*., 1999; Gardner and Davies, 2007). This problem further became more complex as many non-native English speakers decided to avoid utilizing phrasal verbs altogether especially those intermediate level of proficiency learners and also the beginners. Native language speakers also avoid using that forms of verbs during communication (Gardner and Davies, 2007).

Due to the great difficulty of phrasal verbs presented to language learners especially those who learn English as a second language or as a foreign language, the extremely high frequency of English language requires more studies on phrasal verbs and the fact that most beginners and intermediate level of English proficiency avoid using phrasal verbs are a huge problem to English language acquisition. Most reason of avoidance was the complexity of semantic structures and syntactic structures of phrasal verbs and cross-linguistics distinctions which makes most learners to be overwhelmed by enormous number of phrasal verbs. However, linguist must further investigate phrasal

University of Malaya

verbs applications as proven that a very small group of lexical verbs make up a majority of phrasal verbs and as recorded in British National Corpus (BNC) data sources, a total of 100 items in the list of most frequent phrasal verbs, it comprised of top 20 phrasal verbs producing lexical verbs for instance the verbs *take*, *get*, *go* and *come* (Biber *et al*., 1999; Gardner and Davies, 2007).

2.4 Study of corpus

In the past two decades, many educators and researchers have initiated efforts and attention towards vocabulary in second language education of English language especially in terms of multi-word vocabulary items (Folse, 2004). Huge electronic collections of actual language known as corpora, corpus and singular with robust software and high powered computers provide advanced classification and identification of English elusive structures that may also be used to permeate other languages as well (Stubbs, 2007; Granger and Paquot, 2008). However, it is clear that the surface of this complex issue has barely been scratched although corpus linguists have the capability to give more information related to this matter (Read, 2004).

One of the most essential element of corpus study is frequency, that is a study of corpus lean towards quantitative studies or more accurately comparative frequency (Hunston, 2006). Comparative frequency analysis can be performed in many ways, comparing spoken and written word frequency or comparing frequencies of word usage among that of native and non-native speakers, which is quite favoured among corpus linguists in recent times (Wang & Shaw, 2008; Siyanova & Schmitt, 2008; Chen & Baker, 2010, Granger & Paquot, 2015).

University of Malaya

Besides frequency, word processing softwares such as Wordsmith Tool and AntConc are equipped with concordancing tools that enable researchers to examine concordance lines. The study of concordance lines allows researchers to look the context in which searched word is being used and significant co-occurrence of individual element. At this point, researchers will be interested in looking at frequency of co-occurrence, a linguistic phenomenon that create proverbs, idioms, phrasal verbs among others.

The most significant field for advancement and progress in the knowledge gathering on multi-word items is English phrasal verbs and especially the use of lexical verbs in written texts. The study of corpus generates vital insights on recognizing multi-word middle ground between lexis and syntax which is important for applied linguist and implications for second language attainment (Gass and Selinker, 2001).

2.5 Lexical verbs in academic discourse

It is perceptible that drawing out lexical verbs from English for Academic Purposes list of lexical verbs is difficult and frequently fail to provide any signal of word category membership. However, according to the Academic Word List (AWL) by Coxhead (2000), the most popular EAP list are several words that can be verbs and nouns, such as *function*, survey, *approach*, *focus* and *conduct* and the fact that AWL excluded General Service List (GSL), the top 2000 words in English language. The research further justify that academic discourse rarely utilized high frequency verbs in order to express actions, state facts and predicate meaning in a sentence.

Moreover, private verbs for instance verbs like *hope*, *feel*, *like*, *want* and *love* are usually utilized to express personal attitudes, emotions and thoughts, are rarely used in academic discourse but notably appropriate for communication and conversation (Biber,

University of Malaya

1988). However, in some cases of academic discourse, several high frequency verbs have been utilized for academic discourse which turns out to be playing a major role in expressing thoughts and emotions and worthy of being included in EAP syllabuses in which study by Meyer (1997) include verbs like *show* and *find* allowing him to express ambiguities, polysemies and vagueness of daily language but are utilized to elaborate vital part of scholarly process and academic texts. Therefore, Paquot (2007) had included these verbs (absent from Coxhead's AWL) into Academic Keyword List (AKL) verbs such as *suggest, effect, claim, cause, argue* and *aim,* as it is worthy to be covered in academic writing.

Additionally, Swales (2004) stated that a formal research report written in informal English might be regarded as too simplified although the actual data and the ideas are complex. Insufficient exposure of lexical verbs in academic writing is a critical handicap for writers and learners especially for those who learned English as a second language as insufficient exposure or knowledge may prevents them from appropriately expressing their emotions, thoughts and actions in all their shades and expressing them in the expected style (Granger and Paquot, 2015). By exposing the list of lexical verbs or EAP verbs to the learners and writers, their thoughts and emotions and predicate meanings can be conveyed into written text and consequently without a doubt, an imperative first step, but unless it is set off with a detailed depiction of their use, outcomes are bound to be extremely unsatisfactory.

True to the ubiquitous nature of verbs, they have been classified under many labels which largely depend on the purpose that they serve. In relation to academic discourse, five main categories of verbs have been identified. Hinkel (2004) classifies them into the following five categories: activity verbs (*give, take*), reporting verbs (*argue, report*), mental/emotive verbs *(see, write*), linking verbs (*appear,grow*) and logico-semantic

University of Malaya

relationship verbs (*appear, cause*). Certain category of verbs may be dominant in certain area of discourse. In academic writing for example, one may encounter very little emotive verbs and quite abundance of reporting verbs. Justifiably, 'reporting verbs' that has received the most attention in many research papers (Thompson & Yiyun, 1991; Charles; 2006b; Bloch, 2010). However, the five types of verbs are employed in an academic text, regardless of the frequency of use, and therefore need to be studied critically.

2.6 ESL Writers and Use of Lexical Verbs

The development and improvement of academic writing in English as second language learners raises issues of broad significance especially in terms of whether the discourse organization of academic writing in ESL learners and their underlying structures of academic knowledge is either universal or culture specific. The concept of bilingual proficiency must be taken into account along with academic elements of academic discourse throughout cultures and language boundaries. It is perceptible to state that most ESL learners possessed certain level of difficulties in English academic writing especially at sentence-level problems involving vocabulary, grammar and the usage of verb lemmas. Earlier study by Kaplan (1972) found that cross cultural distinctions in rhetoric is the main problem of academic writing by ESL learners and deviations from the expected proficiency of English academic discourse is largely contributed by negative transfer or interference from rhetorical organization of their first language. This subsection will further elaborate on previous studies of the use of verbs in academic writing by ESL writers.

Numerous studies in the field of Linguistic have been dedicated to investigate the perplexing notion of non-native (NNS) discordance from the native (NS) in their writing piece. As mentioned previously, a number of scholars have noted certain quandry that

University of Malaya

NNS writers display in their written products, namely the difficulty to be assertive with the claims and arguing and defending their stance whilst a large number of findings contribute the cause of the phenomenon to be the choice of lexis in their construction of academic prose (Burrough-Boenisch, 2003; Hinkel, 2004). This paper proposes that these two are mutually related.

An academic author needs to be equipped with a repertoire of academic verbs and the knowledge to employ them appropriately in the discourse. One must be able to make use of different reporting verbs *(prove, claim, state*) as they serve different functions, hence colouring the sentences with particular affectations that are dependent to the writers' intention. The writers' ability to manipulate these verbs will provide the readers with correct interpretation of the writers' attitude and stance towards the claims that they made (Manan, 2014). In previous studies, it was found that ESL writers were unable to decide on the proper reporting verbs to state their claims (Pecorari, 2008) and they were often found employing wrong and unnecessary reporting verbs. Bloch (2010) argued that "even if the student can make grammatically correct choices, the rhetorical impact of their claims may suffer if the reporting verb is not appropriate." Hyland (1998) found that one of the major problems that non-native English speakers (NNES) faced when producing academic writing is that they must hold the "definite and self- assured" stance without the sense of "fuzziness". This shows that they do not dare to go beyond their boundary by using "rare" reporting verbs which they seldom use but opt to use the "usual" ones.

An observation that is shared by many scholars is that NNS writers often display limited vocabulary range to the point of repetition in compensation of their limited vocabulary and thinking in L2 whereas their Native counterpart are more prone to variation (Jin, 2008). Hinkel (2004) suggests that NNS writers' texts tend to be built on restricted vocabulary range, to the point where their writing appears repetitive and

University of Malaya

constrained. While it is true that some verbs tend to repeat itself in academic discourse, such as relatively simple verbs such as make, do and look, and some verbs are identified as the common academic verbs (e.g identify, apply, investigate), there are limits as to when it can appear overuse or underuse in an academic prose. Another study on lexical verbs, more specifically reporting verbs at a local university reveals that L2 writers are prone to repeat certain category of verbs (discourse act category) compared to other categories of reporting verb in their theses.

Another frustrating point for NNS writers when they are required to produce an academic piece is the tenses and aspects that exist within the English constructs. Quite a large number of studies have been dedicated to study erroneous uses on tense and aspects in NNS writing and they concluded that indeed this particular area of grammar is considered one of the most problematic (Amsberg, 1984; Ubol, 1981; Kim, 1983; Meziani, 1984; Bryant, 1984; Richards, 1985; Hantrakul, 1990; Pongsiriwet, 2001 as cited in Min, 2013). Typical problems that arise with tense and aspects are inconsistent use, progressive aspects with nonprogressive verbs and the use of passive and active voice (Hinkel, 2004). As the use of tenses and aspects are decoded through the forms of lexical verbs, a two-pronged analysis can be executed on a) the choice lexical verbs and b) verb forms where choice of tenses and aspects can be analysed.

Numerous studies in the field of Linguistic have been dedicated to investigate this perplexing notion and majority of the findings contribute the cause of the phenomenon to be the choice of lexis in their construction of academic prose (Burrough-Boenisch, 2003; Hinkel, 2004). Hinkel (2004) suggests that ESL writers' texts tend to be built on restricted vocabulary range, to the point where their writing appear repetitive and constrained. While it is true that some verbs tend to repeat itself in academic discourse, such as relatively simple verbs like *make, do* and *look,* and some verbs are identified as the

University of Malaya

common academic verbs for instance *identify, apply, investigate* there are limits as to when it can appear overuse or underuse in an academic prose.

Lexical verbs in academic discourse create a minefield of difficulties for writers and enable the writers to modulate the information through voice, mood, aspect and tense conveying into academic texts (Hinkel, 2002; Swales and Feak, 2004). Many studies have been conducted to highlight these difficulties particularly in terms of tense and aspect, and also the issue on transferability of General English rules to English for academic writing (Swales, 1990). Nevertheless, the issue of difficulties and transferability are not the only problem faced by writers in academic discourse as they also have to deal with the fact that EAP verbs required its own lexico-grammatical company, adverbs (such as; *generally accepted, closely related, vary widely, widely used, apply equally, differ significantly*), objects (such as; *provide information, provide evidence, support hypothesis, support the view*), viz subjects (such as; *the evidence shows that, this study suggest that, these outcomes suggest that*) and it also most likely to be use in routinized structures (such as; *it should be noted that, there is little evidence that, as discussed in*) (Granger and Paquot, 2015). Generalities for instance 'the passive is very frequent in academic discourse' are not very accommodating as certain lexical verbs used in academic writing are scarcely ever utilized in the passive while others if not entirely, normally used in the passive (Swales, 2004).

Furthermore, lexico-grammatical restrictions in academic discourse are usually excluded in academic text in which normally present verbs independently from adverbs and nouns when in reality, the interaction between nouns and adverbs caused problems for learners (English as second language) as showed by recent analysis of learner corpus-based studies. Nesselhauf (2005) supported this statement because in his study that involved German-speaking English as a second language, most students misunderstood

University of Malaya

the collocations in the combinations of verbs and nouns and congruently, Hyland (2008) stated that Cantonese-speaking English as a second language, on a study about the word clusters in academic writing, he found that most of the word clusters used in published academic writing are rarely used in the learners' academic writing.

It is wise to say that most of the studies showed that most learners of English as a second language found most difficult on viz that include both much looser routinised sequences and highly fixed routinised sequences, it is phraseology in the broad sense. Few of these phraseological difficulties are shared among novice native writers especially those related to discourse patterns and pragmatic appropriacy (Granger and Paquot, 2015). In a study by Hyland and Milton (1997), novice native writers and Cantonese-speaking English as a second language, usually mixed formal written forms and informal spoken forms and transfer conversational uses of academic discourses and congruently, Neff *et al*. (2004) in their study conducted comparison analysis on the expression of writer stance within various corpora of argumentative texts among professional native writers, novice native writers and English as a second language learners, and the study found that native learners and non-native learners possessed the novice-writer feature of unnecessary visibility. However, we might underestimate the issue if we concluded that English as a second language learners and native student writers encounter the same difficulties in academic writing and can for that reason be regarded as belonging to one and the similar category of novice authors.

2.7 Verb Lemmas vs. Verb Forms

Generally, the fundamental function of verb lemma in corpus linguistic is to allow the corpus linguist to take a broad view of groups of words which in some cases their individual distinctions are irrelevant for instance in examining collocations or compiling

University of Malaya

a word frequency list, we might concordance on a noun irrespective of whether it is in singular or the plural and we might want to treat all the verb forms together. These groups of words could be created and used when it is necessary such as by utilizing regular expressions (Knowles and Zuraidah, 2004). In the practice of corpus linguistic, inflection is the first category to be ignores as inflectional morphology largely regarded on grammatical class and the fact that verb lemma is known to be tied to conventional parts of speech, in case where we talk of the declension of pronouns and nouns and perhaps adjectives and also the conjugation of verbs (Knowles and Zuraidah, 2004).

According to Francis and Kucera (1982), lemma can be described as a set of lexical forms that possessed the similar stem and originated to the similar major word class with distinction only in spelling or inflection. However, this definition of verb lemma raises issues in terms of principle but we could not ignore the fact that it is essential in practice. The constraint to verb forms with the similar stem garner issues especially on what should we do with suppletive forms for instance *went* or *go* and the verb forms of *be*, and if the verb lemmas is *see* the verb forms are *see*, *saw*, *seen*. A criterion of a different order has been introduced by spelling variants where it is one thing to cluster dissimilar representations of what counts as exactly the identical linguistic item and it is another to cluster singular and plural as unique and different, but interrelated linguistic items (Knowles and Zuraidah, 2004).

Moreover, apparent connection can be observed between what lexicographers accomplish when they cluster word forms under headwords and what corpus linguists accomplish when clustering words under a single lemma. In a study by Crystal (1997), lemma has been defined as a dictionary headword or an abstract representation which further subsume all the formal lexical variations that may apply to the headword for instance the verb *walk* is subsume into *walks*, *walking* and *walked* but it is not all

University of Malaya

understandable on the distinctions between a lexeme and lemma. Some researchers argue that defining lemma as a headword is certainly not meticulously definite theoretical concept because clustering words under headwords varies depend on the needs of the intended users and the size of the dictionary and also according to the need of a particular writer on what he decided to do about it. In some study for instance by Kennedy (1998), headwords and verb lemmas are treated similarly as it is typical to lemma the inflectional variants or to list under the same headword in corpus studies.

In terms of assigning group of words to lemmas and clustering words under headwords, researchers recognized the boundary between homonymy and polysemy that will affect the clustering of words for instance the metaphorical and concrete uses of *crane* (which may be refers to machine that lift heavy objects or types of bird) are most probably to be considered as independent words and elements of different lemmas while the metaphorical use of *lion* is probably to be considered as the same word. If it is not easy to cluster word meanings under headwords at the abstract level of the dictionary, it will be much more complex to cluster words in text unequivocally to their lemmas (Crystal, 1997; Kennedy, 1998; Knowles and Zuraidah, 2004).

In a study by Biber *et al*. (1998), they utilized verb lemma casually to regard as the diverse forms of the word jointly with regard to frequency lists and they describe verb lemma rather loosely as the base form of a word which pays no attention to grammatical changes for instance plurality and tense. They further utilize small capitals to symbolize the verb lemma such as DEAL for *deal*, *deals*, *dealt* and *dealing*, which is a usual way of representing the name of a set of words and consider the small capital but highly significant, further step to describe the lemma as the name of a lexical set for example DEAL = {*deal*, *deals*, *dealt*, *dealing*} (Kennedy, 1998; Knowles and Zuraidah, 2004).

University of Malaya

Furthermore, in any corpus-driven studies of lexical verbs, it is important to think about the advantages and disadvantages of using verb lemmas or verb forms as units of scrutiny. If verb lemmas are used in the analysis therefore the different inflectional forms are merged for example *claim, claims, claiming, claimed* because it is a functional option if the aim of the study is to perceive patterns of use that intersect verb forms (for instance the use of a *that*-clause with the lemma CLAIM) and/or to provide a broad overview of learners' lexical repertoire. On the other hand, Sinclair (1991) pointed out that lemmas are a generalization and an abstraction and merely using lemmas amounts to losing imperative information as every verb form has its unique individual patterning and he notices a potential for a new division of study that centralized on the inter-relationships of a lemma and its forms as it is not until now understood how meanings are dispensed among forms of a lemma. Sinclair (1991) suggested that lexicographers alter the conventional practice of utilizing the uninflected form or base as headword and instead, make use of the most frequently encountered form, a revolutionary view that has so far gone unheeded (Granger and Paquot, 2008).

Granger and Paquot (2005) previously conducted an automatic assessment between a similar-sized fiction corpus and a one-million-word corpus of academic writing by utilizing the criteria of frequency, keyness, range and evenness of distribution which then they identified 930 lexical items that figured more significantly in the academic corpus compared to in the fiction corpus. Other than that, the study found out that verbs frequently function as academic writing keywords in merely one or two inflectional forms as they illustrate in their results, approximately 47% or nearly half of the verbs emerge as unique academic discourse items in only one-word form and approximately 23% or almost a quarter of them in two word forms. A minority emerge in three or four or five word forms, respectively 19% and 11% which illustrate that the verb lemma ASSOCIATE, BASE, CONFINE and LINK, emerge as unique academic writing item in only one-word form

University of Malaya

which are *associated*, *based*, *confined* and *linked* (the –ed form). In a case where the verb lemma is LACK or COMPRISE, the verb forms is *lacking* and *comprising* (the –ing form) that is unique and for REVEAL and ENTAIL, the verb forms is *reveals* and *entails* (the –s form). This study explains that we need to be wary about claiming generality and simplification for families whose collocation environments and meanings may be at variance transversely, each inflected and derived word form, and this notion has been highlighted in many studies of lexical verbs in academic writing (Oakey, 2005; Hyland and Tse, 2007).

In addition, verb lemmas had been treated similarly as headwords and as cluster of inflectional variants in a study by Leech *et al.* (2001), but verb lemmas are represented as sets of lexical items in the main text with their members of verb forms listed in italics and they show differentiation between the simplex form and the lemma. The typical practice of lemmatizing of *dealing* to *deal* leaves the theoretical standing of *deal* unclear and indefinite which the only reasonable and logical explanation for this matter is to consider *dealing* as a member of the set DEAL. Consequently, this explanation leads to the fact that the lemma DEAL and the simplex form *deal* are logically distinct from one another and *deal* must be considered as a member of DEAL. Leech *et al.* (2001) is considered to be inconsistent in terms of using sets and spelling variants in a commonsensical manner (a member of lemma can be composed as a set of spelling variants) by listing *realize* and *realise* as separate lemmas but treating *focused* and *focussed* as members of verb *focus* (Knowles and Zuraidah, 2004). These diverse approaches formed what might be known as the traditional view of the lemma, assuming the lemma as a group of words that for realistic functions can be considered as variations of the same word. Conversely, it has turn out to be perceptible that individual members of the lemma can be treated separately and build up their own collocations and meanings, as linguists have started to scrutinize increasingly large corpora for instance *provided* is

University of Malaya

the top participle of the verb *provide* but it has taken on entirely new function as a subordinating conjunction (Leech *et al*., 2001; Knowles and Zuraidah, 2004; Oakey, 2005; Hyland and Tse, 2007; Granger and Paquot, 2008).

Congruently, the verb can still be considered and clustered as the same member of lemma however from a reserved point of view, linguists have to be cautious about what conjectures and presumptions are drawn from lemma membership and linguists without doubt should make undeviating conjectures about meaning or distribution. According to Sinclair (1987) which gathers major insights from several studies of the Birmingham school, as linguists scrutinize further data and focuses on more detail, they tend to generalize lemmas in a much less convincing way and they tend to not treated individual words meaning appropriately (Leech *et al*., 2001; Knowles and Zuraidah, 2004). A study by Stubbs (1996) elaborates on the collocations of words related to *educate* in which the most common word for educate is *education*, frequently related to *higher education* or institutions while the second most common word is *educated*, which habitually related with *at* and the name of a esteemed and influential institution for example *educated at Cambridge*. At this point, whether or not *education* and *educated* fits in the same group of lemmas it is no longer an important issue.

In another study by Tognini-Bonelli (2001), she challenges the postulation that members of a lemma are bound to share similar meaning and vary only in their grammatical profile and she further highlights the application of words *faced* and *facing*, the former having only the metaphorical connotation for instance *faced with a dilemma* while the latter having a tangible connotation for instance *facing forwards* as well as to the metaphorical connotation of *facing a dilemma*). Besides, the solid implication of *facing* occurs in what she recognized as general English or also known as the Birmingham corpus, but not recognized in either Wall Street Journal corpus or the more specialized

University of Malaya

Economist Journal corpus. Therefore, it is complex to see beyond pronunciation and spelling and the significant of generalizations that might be captured by assigning *faced* and *facing* to the lemma word FACE (Leech *et al*., 2001; Knowles and Zuraidah, 2004; Oakey, 2005; Hyland and Tse, 2007; Granger and Paquot, 2008). Nowadays, the term lemma is applied to describe a number of concepts which might be inter-related with distinguish logic: spelling variants, the name of a set of lexical items, a set of paradigms, a set of inflectional variants, a dictionary headword and an ad hoc group of words. The notion of the lemma is mainly practical at a common level in highly abstract elaborations of a language, although it seems to be of uncertain value for comprehensive studies of real texts (Granger and Paquot, 2015).

2.8 Language Formulaicity

Language formulaicity, or phraseology is suggested by Wray (2002) to be pervasive in all language data. Upon his observation of daily language use, Sinclair (1991) commented that the majority of it consists of recycling common words in similar patterns and that most of these words actually lack exclusive meaning. He added that they can only be meaningful when used with a selection of words, with which they often appear together in a broader repertoire. During his study of academic writing consists of approximately 238,000 words, Howarth (1998) noted that 31–40% of the essays contained collocations and idioms and Altenberg (1998), in his analysis of the London-Lund Corpus, reported that approximately 80% of the words that made up the corpus formed part of "recurrent word combinations". A study by Conklin and Schmitt (2008) further indicates that collocations are indeed pervasive in language discourse and the diference in level of pervasiveness is that differentiate the speech and writing of native and non-native speakers. Such finding is supported by Erman and Warren (2000) study

University of Malaya

who analysed native speakers' written and spoken discourses. They found out that formulaic expressions or collocations make up 58.6% of the spoken English discourse and 52.3% of the written discourse. Additionally, Foster (2001, as cited in Conklin and Schmitt, 2008), in their analysis of formulaic language in informal natives' speech, reported that 32.3% of speech consists of formulaic expressions where the non-native counterpart makes significantly less formulaic language during a similar context.

A study of a science corpus also yields large number of collocations. A study of science corpus by Sujatha Menon & Jayakaran Mukundan (2012) have discovered a high percentage of Noun/noun and Adjective/Noun collocation and concludes with a suggestion of the importance of language learning in chuncks rather than individual words. The conclusions to which all these studies have arrived to is that formulaic language forms a large part of any discourse (Conklin & Schmitt, 2008) and that knowing these formulaic sequences can facilitate the process of language learning. Proponents of collocation agree that theory of formulaic language disputes Chomsky's claim that language production is a description of the speakers' and hearers' intrinsic competence (1965). The significance of language formulaicity is that it economised the language by subjecting itself to the basic of language principal but simultaneously remain simplistic in nature. Consequently, language production is no more reflective of the speakers' or hearers' actual language competence.

Due to the significantly ubiquitous nature of collocation, a large number of studies have concurred to the suggestion that the frequency-based language formulaicity which are largely found in native writing can be of great help to non-native writers to achieve a more native-like style of academic writing, and should thus be integrated into ESL/EFL curricula. A study by Ellis et.al (2008) further reaffirms the benefits of formulaic sequences whereby he also advocates the use of academic language formulas in EAP

University of Malaya

instructions. In his study, he triangulates data from three different perspectives; psycholinguistics, education and ESL. Through that study, he successfully proves that the study of formulaic language is relevant from all three perspectives.

In the numerous literature this paper has reviewed, one article is particularly cautious of the use of of formulaic language. In his thesis, Kjellmer (1991, as cited in Nesselhauf, 2005) proposed that the argument that learning and teaching of formulaic language can contribute to native-like writing is a persuasive one and he suggested that researchers as well as learner need to treat the hype with slight caution. Kjellmer's point is more often than not, misundertood to have implied that learners will sound more like their native counterpart if they use formulaic language liberally in the ir speech and writing. However, that is not always the case. A study by Durrant & Schmitt (2010) found that although there are non-native writers who utilise high-frequency collocations, they still do not sound native-like at all. This is because while they use high-frequency collocations, they chose to underuse less frequent, strongly associated collocations, items which are probably highly salient for native speakers. These findings are consistent with "usage-based models of acquisition" by Durrant & Schmiddt (2009) while accounting for the impression that non-native writing, lacks language formulaicity.

Similar observation is made by Sujatha Menon & Jayakaran Mukundan (2012) in their study of Science textbook in Malaysia, They strongly propose that "these patterns should not be over-represented". Learner should only be taught of the frequent collocation patterns but need to be cautioned of the changeability and diversity of collocational patterns in scientific language. They note that that the flexibility of some patterns can be "arbitrarily blocked by usage" and that these "arbitrary lexical patterning" pose a large hindrance for learners. However, these studies ultimately reveal the general consensus

University of Malaya

that learners who does not possess adequate knowledge of formulaic language will equally possess lower language proficiency and fluency.

Considering the ubiquitous nature of language formulaicity based on previous literatures, one should think that ESL textbooks should also provide authentic language exposure that it similar to that of native environment. Surprisingly, however, this is not always the case. A number of textbook evaluation studies have shown that instead of teaching authentic language, they show the learner how the language should is used in formal contexts, not at all reflective of language use in authentic settings. Language use in authentic setting is often marked by frequent use of phrasal verbs, as they are often used by native speakers. A study of a national textbook by Faizah (2013) to examine the usage of phrasal verb in Year Three English Textbook published by the Dewan Bahasa & Pustaka under the permission by Ministry of Education Malaysia for the new curriculum, Standard Curriculum for Primary School yield dissapointing results. The instances of phrasal verbs are identified and the frequency of usage are recorded and compared to the list of 50 Most Common Phrasal Verb. The results will indicate whether Malaysian Year Three English Textbook primes young learners to the natural and authentic use of the English Language. Out of the 28 phrasal verbs, only 10, or 36% of them are repeated while the others are mentioned only one. These results reflected the findings of a study by Darwin and Gray (1999) who perform a comparison between a list of 20 most common phrasal verb of the BNC (British National Corpus) to that of a typical ESL exercise book. They found that only three of all the phrases in the textbooks matched the 20 phrasal verbs on the list. It would seem like the ESL society have not caught up with these findings and the situation should be remedied soon.

University of Malaya

2.9 Language Formulaicity; Phraseology and Collocation

Language knowledge is collocational knowledge (Gitsaki, 1996). The term phraseology and formulaic language or formulaic sequences (Granger and Meunier, 2008; Schmitt, 2008) are often used to elaborate the various kind of multi-words units and recent research shows studies using corpus data in order to highlight the significant of multi-word units in language had increase in these past few years. Altenberg (1998) in his study estimated approximately 80% of the words in the corpus produced part of intermittent word combinations while Wray (2002) found that learners face difficulties in terminology when explaining word co-occurrence. It also appears that these terms; language formulaicity, phraseology and collocation can and have been used interchangeably at large. The term phraseology can be used in distinctive ways by various scholars but various terminologies can be used to interpret the same notion or similar word of co-occurrence for instance the term *recurrent word combinations* (Altenberg, 1998), *clusters* (Hyland, 2008), *n-grams* (Stubbs, 2007), *phrasicon* (De Cock *et al*., 1998) and *lexical bundles* (Biber and Barbieri, 2007). These terminologies (*recurrent word combinations, clusters, n-grams, phrasicon* and *lexical bundles*) actually predicates to continuous sequences of words acquired through corpus-driven method with distribution criteria and specified frequency and the sequences are utilized by the native language within specified contexts and are fixed multi-word units that possessed customary discourse functions or pragmatic functions (Granger and Paquot, 2008; Chen and Baker, 2010).

In a chain of lexical bundle research by Biber *et al*.(1999), Biber *et al*. (2004) and Biber and Barbieri (2007), academic prose and conversation present different distribution patterns of lexical bundles in which most lexical bundles in academic prose are phrasal while lexical bundles in conversation are clausal. Cortes (2002) studies lexical bundles in native freshman compositions and stated that novice writers utilized lexical bundles in a

University of Malaya

different functional compared to published academic prose and in Cortes (2004) he found that native students rarely apply lexical bundles in academic writing and even if native students applied lexical bundles, it is applied in a different manner. Non-native competence in terms of collocational use can be observed from a number of sources. Majority of previous studies have analysed real-life language production, in either written or spoken forms. Computerised analysis has enabled researchers to explore all aspects of phraseology occurring in a large body of texts.

It has been widely concurred by researchers and language teachers alike that phraseological knowledge plays a pivotal role in language learning (Nation, 2001). They agreed that the "appropriate use of collocations enables the learners to speak more fluently, makes their speech more comprehensible and helps them produce more native-like utterances" (Manan & Noor, 2014, p.2) and consequently is deemed profound in the process of Second Language Acquisition (SLA) (Durrant & Schmitt, 2010). Cortes (2004), for instance has noted that the use of collocations and fixed expressions has been considered a marker of proficient language use, and agrees with Haswell (1991, p.236) report that "as writers mature they rely more and more on collocations". Similarly, more recent studies by Nesselhauf (2005) and Kazsubki (2000) in which they explore the development of collocational knowledge in non-native writing have come into conclusion that there exists correlation between assumed increased proficiency with augmented use of common phraseology. In other words, it can be assumed that as the learner develop their linguistic proficiency, the frequency by which they use conventiona; collocations will increase as well.

It is established that native speakers share a substantial body of formulaic sequences thus it is the second language speaker's ability to gain phraseological knowledge that determines, in part, language learning success. Bearing the learners' as

University of Malaya

well as instructors' needs in mind, researchers have been looking this aspect of language that could help non-natives achieve similar competency as their native counterparts. Granted, achieving sufficient phraseological knowledge is not as simple as we would like to imagine. It is suggested that Language environment is what separates non-native speakers' ability in utilising phraseology to their native counterpart (Wray, 2002). Native speakers are able to apply the collocational knowledge easily as they recognised the formulas as unit with clearly defined functions, due to vast exposure of the target language (Ellis et al., 2008). Contrarily, non-native speakers, less exposed to the language, flounders in using phraseology. Non-native can only absorb this knowledge if they were immersed in a speaking community where they can observe and imitate (Wray, 2002). Siyanova and Schmitt (2011) have conducted a study of the influence of native-speaking environment to non-native production of collocation and found that those who have spent more time in such environment are more inclined to use collocation. This finding, however striking does not provide a solution to the phraseological conundrum. Not everyone can manage the opportunity to spend extended time in native environment and the dilemma remains unresolved.

Considering the previous discussion, it can be observed that different researchers have adopted specific terms to refer to phraseology and to determine the most accurate terminology to use as an umbrella term is proved a very difficult endeavor. It is also noted that there has been very little effort contributed to achieve unified definition and criteria of select terms. In the face of this chaos, Gries (2008) attempts to clarify the term phraseology in effort to make it applicable to research in the area co-occurrence phenomena in the field of linguistics. Gries (2008) chooses the term phraseology as the umbrella term for word combination as it covers a very broad concept therefore allows present and future research to be influenced and identified with the terminology. He characterises phraseology according to three approaches to linguistics. In relation to this

University of Malaya

study, the six parameters applied to characterise phraseology in the corpus linguistic paradigm are discussed. He noted that other terms which are also widely used in this paradigm are collocation, n-gram and cluster.

Table 2.1 Parameters of phraseology in corpus linguistics

nature of the elements	words
number of elements	*n* (usually, that means 'two or more')
frequency of occurrence	sufficiently frequent to be recognized as a combined element
distance of elements	for clusters/*n*-grams, the distance is usually 0 (i.e., the elements are immediately adjacent); for collocations, the distance between the elements involved can vary, but usually exhibits one or a few preferred distances
flexibility of the elements	for clusters/*n*-grams, there is usually no flexibility; for collocations, one usually allows for some flexibility: the collocation of *strong* and *tea* would be instantiated both by *strong tea* or *the tea is strong*
semantics	*n*-grams are usually retrieved for natural language processing purposes where the issue of non-compositional semantics is only sometimes relevant; for collocations, researchers differ as to whether they require some non-predictable behavior (*strong tea* is acceptable but *powerful tea* is not) or not.

Source: Gries , 2008 (p. 16)

University of Malaya

2.10 Summary

This chapter has discussed at length various literature concerning academic genres, verbs in academic writing and ESL learners in their use of lexical verbs in academic prose. It is generally agreed that ESL writers face a challenge in composing a sound and cohesive piece of academic writing due to several reasons, mainly in choosing appropriate verbs to enhance their statements and claims and also lack of exposure to academic lexical bundles or stock phrases. There are also valid concerns regarding L1 transference to their L2 writing although these seem to affect novice writers more than the expert writers. The suggested solution to these conundrums appears to be by teaching or exposing learners to the frequently used phrases related to their genre. However, these matters need to be studied further with focus on verb usage, tense and aspect and phraseology in a collection of theses by advanced ESL writers. The methodology employed in this study will be discussed in further detail in Chapter 3.

University of Malaya

CHAPTER 3: METHODOLOGY

3.1 Introduction

This chapter describes in detail the procedure of collecting and analysing the data. The rationales of the procedures are also explained. First, the methodology chosen for the present study is described and discussed, followed by the method for corpus selection, corpus size, method of analysis and the software used to assist analysis.

This study intends to describe features of lexical verbs by identifying categories of verbs used in the ESL corpus gathered. The examination of verbs forms could also provide a deeper analysis of tenses and aspects, voice and associated patterns favoured by the writers and common associated patterns of lexical verbs are also identified and examined. Phraseology are a lexical phenomenon that has linguistic and lexicographic status as well as utility for statistical natural language paradigms (McKeown & Redev, 2006). That is to say, phraseology itself is the result of language use and continuous research. In order to process large amount of data, researchers are now able rely on computational linguistics to extract patterns of phraseology from text corpora.

3.2 Analysis of authentic language data

Data on the phraseological aspects of non-native competence can be derived from various sources as there are several ways of directly investigating the use of collocations by learners utilised by previous studies. More often than not, it can be obtained by analysing the language production of learners, either written or spoken. A growing number of second language learner (L2) corpora have emerged in recent years; for example, the CALE (The Corpus of Academic Learners English) and the CLEC (Chinese Learner English Corpus) and our very own The Malaysian Corpus of Learner English (MACLE) as interest in this field continues to sustain.

University of Malaya

In comprehensive measures, particular samples of L2 writing or speech are analysed with respect to all the collocations which occur in the available texts. Another method is to use corpora of non-native writing, in which case it is possible to analyse only specific, pre-selected collocations as they occur in a range of texts as done in a study by Granger (1998). By obtaining concordances for the investigated items, the collocational patterns of non-native texts can be compared to those in texts produced by native speakers. The present study attempts to gather data through the mean of corpora based on advanced non-native academic writing. This method is referred to as corpus linguistics.

There have been some serious debates over the term corpus linguistics to this day a solid definition has yet to be concluded. One of the earliest definitions came from Leech (1992) in which Corpus Linguistics is labelled as a new paradigm for computerized linguistic research. Corpus linguistics can be defined as "…the study of language based on examples of 'real life language use" (McEnery and Wilson, 2001, p.1). Other researchers have defined corpus linguistics as "[...) a way of investigating language by observing large amounts of naturally-occurring, electronically-stored discourse, using software which selects, sorts, matches, counts and calculates." (Hunston and Francis, 2000, p. 14). As we can deduce from the definition, corpus linguistics in itself is not a branch of linguistics, like syntax and semantics. Rather, it is a methodology and technique for language study that can be used in any branch of linguistics. However, further arguments emerge in the face of diversity in practice of this so-called methodology. More recent argument is posed by McEnery and Gabrielotos (2006) in which they claim that the differences in practice of corpus linguistics are underlined by theoretical considerations as Teubert (2005, p.2) describes Corpus Linguistics as "a theoretical approach to the study of language".

University of Malaya

3.3 Corpus Linguistics

Corpus linguistics have been considered as a theory by many linguistic researchers (Teubert, 2005), whilst some consider corpus linguistics as a methodology (Bowker and Pearson, 2002). The more widely accepted definition of corpus linguistics, however, is by McEnery *et al*. (2006) that describe corpus linguistics as a comprehensive system of techniques and principles on the approach to corpora application in language studies and language teaching and learning, and corpora certainly possessed theoretical status yet the theory is not in itself. Teubert (2005) elaborates corpus as a discipline or methodological commitment that is rather an assertion on being used with only real language information taken from discourse in a principled and standardized way and later assembled into a corpus.

Corpus-driven linguistics does not take traditional linguistics for granted but making full use of it as a discourse and not a language-peripheral taxonomy of linguistic units which will have to present the classifications and categories that are required to answer a prearranged research question (Teubert, 2005). On the other hand, corpus-based linguistics approached corpus data from the standpoint of moderate corpus-peripheral premises with the objective of improving, modifying and testing such theories and frequently utilized corpus annotation (Hardie & McEnery, 2010).

Corpus linguistics is not merely a simple recondite field of study within linguistics but it requires practical tools and methodologies to further analysed various aspects of language use, not least in areas related to learning and teaching foreign language or second language (Bolton & Tyne, 2013). The impact of corpora within linguistics study has been described as revolutionary as it facilitates the field of linguistic to show how language should be utilized in authentic communicative occurrences and have been influential in informing course materials, evaluation approaches, syllabuses, usage

University of Malaya

manuals, books, grammar and dictionaries. Corpora can also be directly utilized for data-driven learning applied by learners and teachers, and especially those who believe that corpora provide learners with total control of how to look and what to look for, with one thing leads to another in an unintended manner (Bolton & Tyne, 2013).

Corpus can be described as a notion that compose of a radial characteristic of the similar kind as a polysemous word and constitute of prototypical exemplars by virtue of exhibiting few extensively acceptable characters and also constitute various exemplars that are less directly or related to the prototype, to other exemplars of the grouping by family similitude links (Cox, 2011). Studies by Bhatia *et al*. (2011) mentioned that corpus comprised of several characteristics which are one or more machine-readable Unicode text files, representative of specific type of language or variety or register or speaker as a whole with the sampling scheme in corpus analysis should indicates population variation that the analysis should represents. Other than that, corpus linguistics possessed character of balanced because corpus requires the sample size to be proportional in the study population, and it must contain information from natural communicative settings which refers to the purpose of producing language data in the corpus must be untainted by the collection of those data (Biber and Conrad, 2009).

Nowadays, the field of corpus linguistics has expanded into many traditional field of language analysis. For instance; sociolinguistic variation, diachronic linguistics and indeed, ESP. Seminal work remains as a theoretical reference of primary significance and the utilization of corpora in linguistics studies and applied linguistics in the modern days does not require defense but insight, proficient research skills and comprehensive method (Benjamins, 2014).

University of Malaya

3.4 The corpus

The general principle of collecting a learners' corpus is that selection and compilation process needs to be made according to strict design criteria that controls the variables regarding the learner (e.g. gender, age, language background) (Dagneaux et al.,1998). This is because even in a homogenous language background such as ESL, a number of concerns involving other variables need to be addressed. This study concentrates on corpus-driven investigation of submitted Master's theses to Faculty of Language and Linguistics. Thirty-five Master's theses are selected for this study, whereby the authors are considered to be advanced ESL writers of English due to the University's requirements; undergone English medium undergraduate courses and a bachelor's degree with CGPA of not lower than 3.0 or equivalent. The courses are also fully conducted in English. These theses were accepted by the faculty and the writers were awarded Master of Arts between year 2010 to 2014 as part of the requirements for the degree of Master of English as the Second Language or Master of Linguistics. Additionally, these theses were uploaded onto the university's digital library for easy access. Only Discussion sections (marked by subheading Discussion) are analyzed in this study. A whole discussion chapter would be a much ideal data for this study, however the majority of the theses are not compiled in that manner. The discussion is normally placed in Chapter 5 which takes only a section of that chapter, hence the data is collected as such. The population is also determined to be of ESL background by their Malaysian Identification Card number. The summary of the corpus is presented in Table 3.1.

Table 3.1 Summary of corpus

Type	Number of text	Years	Total Words
Master of Arts dissertations (Discussion section)	35	2010-2014	55 252

University of Malaya

3.5 Data Collection Procedures

The first step for data collection was to browse University Malaya's repository website at http://www.diglib.um.edu.my/umtheses/ where electronic versions of theses and dissertations from all faculties are uploaded and stored. From there, the public are able to choose the faculty and download PDF formats of all the dissertations. Dissertations are chosen based on these criteria;

1. The writers must be a Malaysian citizen to fulfill the characteristic of ESL background. Identification card number is considered as evidence for this purpose.
2. The dissertation must have a heading or subheading Discussion in either Chapter 4 or 5 where the candidates discuss their findings of the study.

3.6 Ethics of the study

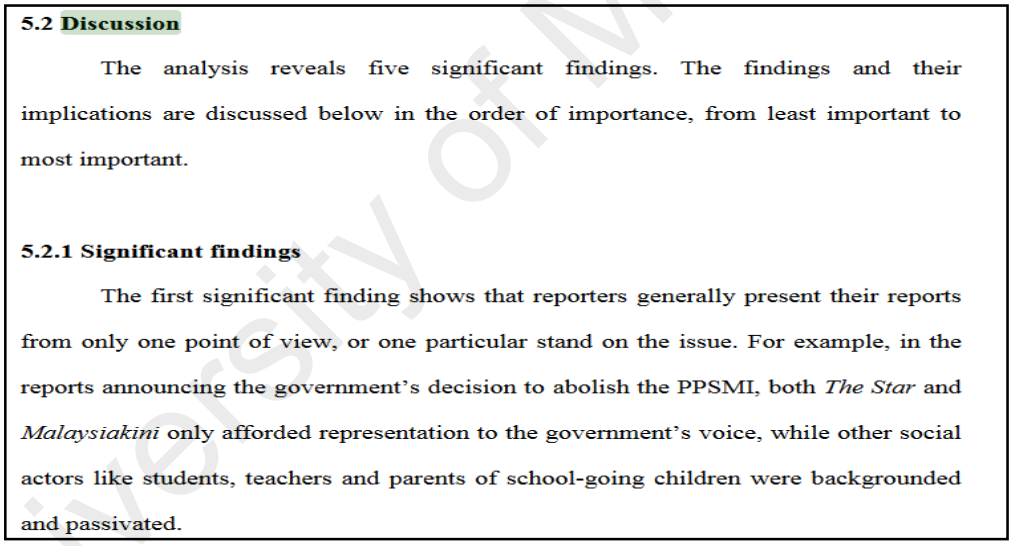
While ethical issues are deemed highly relevant to corpus linguistics as with any other branch of linguistics, McEnery & Hardie (2012) note that the literature in Corpus Linguistics has paid little consideration to this aspect. They point out that reference textbooks related to the discipline written by the likes of Sinclair (1991), Biber (1998) and McEnery & Wilson (2001) did very little to address this issue. While the Corpus Linguistics literature is mostly silent on ethical issues, it does generally embody good ethical practice as suggested by McEnery & Hardie (2012). On another note, following the guideline by Lancaster University (2012), if one collects data without seeking permission, distribution is not allowed. This way, there is no copyright breach. This study adheres to these rules of non-distribution, thus preventing copyright breach. In addition, only a few out of context sentences are shown as examples throughout the dissertation, especially in Chapter 4. It is impossible to reconstruct the content of the original text based on these disjointed examples alone. Therefore, even if the content is redistributed,

University of Malaya

it is still not considered as copyright violation. In addition, the study also practices anonymity with regards to the authors of the corpus. The files are labelled in such a way (ESL1-7) avoid usage of the original authors' names.

3.7 Data Analysis Procedures

Each thesis portion is stored in PDF format in the online repository. They have to be converted to .txt format before the data can be analysed with the WordSmith Tools. The software only processes data in .txt format because there are many hidden markups in other formats.



5.2 Discussion

The analysis reveals five significant findings. The findings and their implications are discussed below in the order of importance, from least important to most important.

5.2.1 Significant findings

The first significant finding shows that reporters generally present their reports from only one point of view, or one particular stand on the issue. For example, in the reports announcing the government's decision to abolish the PPSMI, both *The Star* and *Malaysiakini* only afforded representation to the government's voice, while other social actors like students, teachers and parents of school-going children were backgrounded and passivated.

Figure 3.1 Example of text with Discussion subheading

The content marked with subheading Discussion (e.g 5.2 Discussion) is copied and saved as txt. format for further analysis. The txt. format files are collected in 35 separate files and later compiled into 7 files, labelled ESL 1 to ESL 7. The compilation allows for easier and faster tagging process as the software allows for a limit of 100 000 tokens per file. Each file is copied and pasted in the POS tagging software available at **http://ucrel.lancs.ac.uk/claws/trial.html.** Tagset 5 and horizontal output are chosen as

University of Malaya

it has less tags (60) that enables it to process bigger set of data. The tagged files are then processed with WordSmith Tool software.

The_AT0 analysis_NN1 reveals_VVZ five_CRD significant_AJ0 findings_NN2 ._SENT -----_PUN

The_AT0 findings_NN2 and_CJC their_DPS implications_NN2 are_VBB discussed_VVN below_AV0 in_PRP the_AT0 order_NN1 of_PRF importance_NN1 ,_PUN from_PRP least_DT0 important_AJ0 to_PRP most_DT0 important_AJ0 ._SENT -----_PUN

Figure 3.2 Example of text with POS-tags

3.7.1 UCREL CLAWS Tagset

The data is tagged and annoted using UCREL CLAWS Tagset software which is available for free at Lancaster University website. Part-of-speech (POS) tagging is the most commonly used method for corpus annotation. The process in required in order to identify the lexical verbs in the text. This software can expedite the process of identifying verbs in large body of data. CLAW Tagset C5 is utilised considering the large number of token. Similar tagset is also used for the BNS (British National Corpus). The software processes the text in three stages; pre-edit, automatic tag assignment and manual post-edit. Abbreviation forms in relation to verbs for CLAWS are listed in Table 3.2.

Table 3.2 CLAWS Horizontal Output Abbreviation

VVB	base form of lexical verb (except the infinitive)	e.g. SHOW,MAKE
VVD	past tense form of lexical verb	e.g. SHOWED, MADE
VVG	-ing form of lexical verb	e.g.SHOWING,MAKING

University of Malaya

Table continued from page 55

VVI	infinitive of lexical verb	e.g to-SHOW, to-MAKE
VVN	past participle form of lexical verb	e.g. SHOWN, MADE
VVZ	-s form of lexical verb	e.g. SHOWS, MAKES

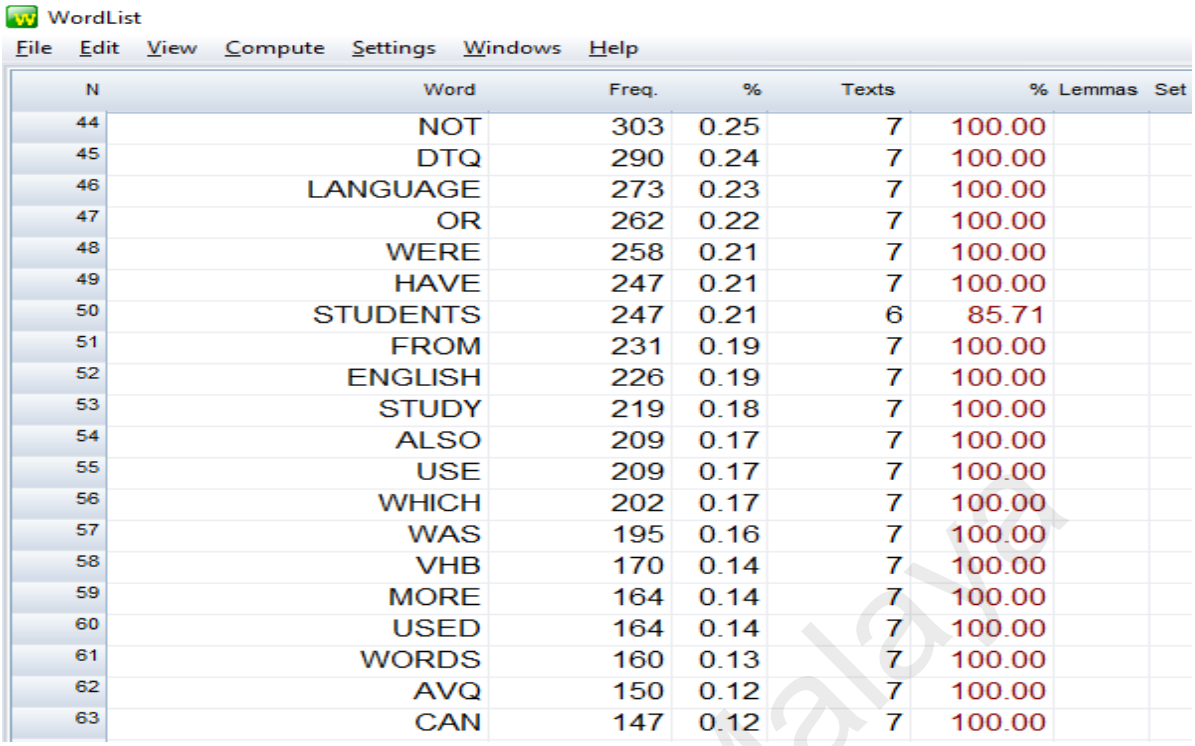
The CLAWS system boasts a success rate up to 97% on written text thus ensures the reliability of the results.

3.7.2 WordSmith Tool 6.0

The text analysis software, WordSmith Tool is developed by Oxford University Press since 1996. The software has undergone a number of upgrades which is now at version 6.0. It is ingeniously developed software, in terms of variety of analysis tools offered; Concordance line generator, Keywords extractor and WordList application. For the purpose of the study, the software is able to generate word lists according to its frequency order, and generate concordance lines to find collocations and show frequencies altogether with statistical tools. It can also compare different texts by showing their statistical significance level. The annoted corpora are later ran through the WordSmith Tool 6.0 to generate wordlist and concordance lines.

The wordlist is created by listing the tokens by frequency. This method enables the study to identify the Top 100 verbs for RQ 1. The complete list of all the tokens is included in Appendix A.

University of Malaya


WordList

File Edit View Compute Settings Windows Help

N	Word	Freq.	%	Texts	%	Lemmas	Set
44	NOT	303	0.25	7	100.00		
45	DTQ	290	0.24	7	100.00		
46	LANGUAGE	273	0.23	7	100.00		
47	OR	262	0.22	7	100.00		
48	WERE	258	0.21	7	100.00		
49	HAVE	247	0.21	7	100.00		
50	STUDENTS	247	0.21	6	85.71		
51	FROM	231	0.19	7	100.00		
52	ENGLISH	226	0.19	7	100.00		
53	STUDY	219	0.18	7	100.00		
54	ALSO	209	0.17	7	100.00		
55	USE	209	0.17	7	100.00		
56	WHICH	202	0.17	7	100.00		
57	WAS	195	0.16	7	100.00		
58	VHB	170	0.14	7	100.00		
59	MORE	164	0.14	7	100.00		
60	USED	164	0.14	7	100.00		
61	WORDS	160	0.13	7	100.00		
62	AVQ	150	0.12	7	100.00		
63	CAN	147	0.12	7	100.00		

Figure 3.3 WordSmith Tool Wordlist

The verbs are identified and listed manually based on the information from the Wordlist. The categories for classification is Hinkel's academic verb types which are activity verbs (*give, take*), reporting verbs, mental/emotive verbs, linking verbs and logico-semantic relationship verbs. The process takes into account the frequency of all forms of verbs (lemma). The list is scanned and analysed for verbs. Some verbs are quiet difficult to categorize. For example, the verb *study* could be categorised as a noun and verb. However, upon further inspection, the word is employed as a noun rather than as a verb as in *the study shows*.., thus eliminating them from the Top 100 list. The process of identifying and eliminating verbs that can also function as a noun is rather a tedious process and takes quite some time to be completed. For example, there are over five hundred entries of study and the author has to go through the concordance lines to examine the context. When the word has to be taken out of the list, another word has to be examined to be included in the list and the process repeats itself. One other major consideration that needs to be made is the frequency of the verbs. In a corpus as relatively

University of Malaya

small is this, the raw frequency itself sometimes does not really seem significant. It can have a frequency of 30 or less and still makes it to the list. The author has considered including normalized frequency into the report but decided against it since this is not a comparative study. Other words such as *write* and *teach* can be either Activity verbs or Report verbs. In this corpus especially, there are an abundance of *write* and *teach* used in particular referring to the act of writing and teaching that occur in a language classroom. They are later categorised as Activity verbs after an examination of their usage in the sentences. During the process of data analysis, a large number of this type of issue is encountered; therefore, great caution needs to be taken during the process of categorisation. A lot of time and manpower has been dedicated into ensuring the list is valid and reliable data.

For RQ 2, the corpus is analysed via concordancing tool. The text is processed through CLAW POS tagger before the concordance lines such as below can emerge. Upon entering VV* into the search bar, a list of verbs labelled of VV* will appear. There are six variation of VV* tags, as shown in table 3.2. The tags are later sorted based on frequency and further analysis on significant verb forms is carried out.

N	Concordance	Set	Tag	Word #	Sent	Sent	Para	Para	Hea	Hea	Sect	Sect	File
535	framework_NN1 has_VHZ enabled_VVN the_AT0 study_NN1			7,057	28	17	0	66			0	66	NNNS1.txt 2017/
536	macro_AJ0 level_NN1 has_VHZ enabled_VVN the_AT0 researcher_NN1			5,831	21	43	0	71			0	71	NNNS7.txt 2017/
537	This_DT0 study_NN1 has_VHZ enabled_VVN the_AT0 researcher_NN1			7,998	28	32	0	96			0	96	NNNS4.txt 2017/
538	has_VHZ been_VBN observed_VVN enacted_VVN through_PRP			7,039	28	62	0	66			0	66	NNNS1.txt 2017/
539	ideologies_NN2 are_VBB also_AV0 enacted_VVN in_PRP other_AJ0			7,797	27	67	0	93			0	93	NNNS4.txt 2017/
540	intellectual_AJ0 perfection_NN1 encapsulated_VVN in_PRP a_AT0			8,590	35	50	0	80			0	80	NNNS1.txt 2017/
541	difficulty_NN1 was_VBD not_XX0 encountered_VVN by_PRP these_DT0			4,639	20	57	0	81			0	81	NNNS2.txt 2017/
542	._SENT -----_PUN This_DT0 has_VHZ encouraged_VVN many_DT0 to_TO0			5,169	18	25	0	63			0	63	NNNS7.txt 2017/
543	prank_NN1 callers_NN2 are_VBB encouraged_VVN to_TO0 never_AV0			4,831	18	50	0	58			0	58	NNNS4.txt 2017/
544	are_VBB highly_AV0 encouraged_VVN to_TO0 be_VBI			2,251	96	42	0	40			0	40	NNNS2.txt 2017/
545	,_PUN participants_NN2 are_VBB engaged_VVN in_PRP the_AT0 93_			3,221	13	50	0	40			0	40	NNNS5.txt 2017/
546	,_PUN the_AT0 participants_NN2 engaged_VVN in_PRP deeper_AJC			3,821	15	24	0	48			0	48	NNNS5.txt 2017/
547	When_AVQ learners_NN2 are_VBB engaged_VVN in_PRP producing_VVG			4,389	17	22	0	55			0	55	NNNS5.txt 2017/
548	,_PUN your_DPS skin_NN1 is_VBZ enhanced_VVN with_PRP natural_AJ0			974	40	93	0	9%			0	9%	NNNS1.txt 2017/
549	the_AT0 teachers_NN2 are_VBB equipped_VVN to_TO0 be_VBI			2,465	10	69	0	23			0	23	NNNS1.txt 2017/

Figure 3.4 Example of concordance lines

University of Malaya

RQ 3 is answered by focusing on verb-based structural forms which are (VP +) *that*-clause fragment passive verb + PP fragment. The process is performed by looking at VV* tags with *that* right collocates and examining VVN tags with PP tags as right collocates respectively.

In summary, the work flow is shown in the Table 3.3.

Table 3.3 Stages of corpus analysis

Stage	Process	Research type
Stage 1	Selection of dissertations from university's repository	Corpus-based approach
Stage 2	Formation of corpora from 35 dissertations	Corpus-based approach
Stage 3	Conversion of corpus into text format	Corpus-based approach
Stage 4	Automatic generation of frequency lists	Descriptive analysis
Stage 5	Identification of verb types and forms	Corpus-based approach
Stage 6	Application of statistical analyses across and within corpora	Corpus-based approach
Stage 7	Analysis of verb patterns	Descriptive and Interpretative

3.8 Summary

In this chapter, we have reviewed the methods for real-life language data analysis which has since branched into an area of study itself called Corpus Linguistics. By employing computational method for data analysis, researchers have been able to explore larger size of data with minimal margin for errors. The process by which the corpus is examined are also described in systematic details. These are the standard procedure for many corpus linguistic analysis. As mentioned earlier, the softwares utilised during the analysis, WordSmith Tool and UCREAL POS-tagger are specifically tailored for linguistic analysis thus ensuring reliability of the results.

University of Malaya

CHAPTER 4: DATA ANALYSIS AND RESULTS

The present chapter has been formulated to present the findings for the current research. It begins by mapping the overall picture of verbs which are categorised into five types of lexical verb- Activity, Report, Mental-Emotive, Linking and Logical-Semantic Verbs. In this part, the study also looks at the verb forms used by ESL writers. This should provide insight into the tenses and lexico-grammatical choices made by the ESL writers. In the second part of analysis, dominant patterns of the verb-based bundles are examined.

4.1 Types of lexical verbs

RQ1: What are the types of lexical verbs used in the discussion sections in M.A theses?

The first aim of the study is to list the lexical verbs following the categorisation established by Hinkel (2004). Identifying the lexical verbs into the categorisation set as the initial step of analysis is important because it will show the functions and purposes of the lexical verbs, whether they are used as report, linking verbs etc. The breakdown of Top 100 lexical verbs found in the corpus can be visually seen in Figure 4.1.

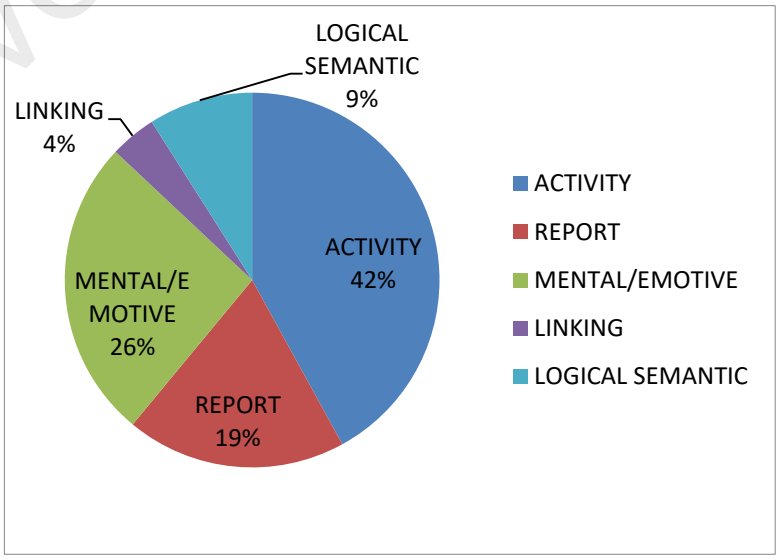


Figure 4.1 Percentage occurrence of lexical verbs in specific verb groups

University of Malaya

Figure 4.1 shows the Top 100 verbs found in the ESL corpus are categorised into five types; Activity, Report, Mental/Emotive, Linking and Logical-Semantic verbs. It shows that the Activity verbs appear with the highest variations compared to other types of verbs (42%, n=42), followed by Mental/Emotive verbs (26%) and Report verbs (19%). The Logical Semantic verbs (9%) and Linking verbs (4%) are used less commonly in the corpus.

The high percentage of activity verb found in the corpus is expected because Activity verbs consists the largest class of verbs that they also include subclass of Activity verbs (Hinkel, 2004). The second most used verbs are the Mental-Emotive verbs. Previous literature suggests that they are scarcely encountered in academic discourse which proves to be a significant contrast to the findings of current study. Further justification on this result is explained later in this chapter. On the other hand, Report verbs are very common in academic writing. The use of Report verbs allows the writers to cite other sources in a precise manner and for making a stance in the argument. As such, they are often identified in a greater number in the Literature Review chapter. This study finds Report verbs as the third most commonly used verb. The least frequent verbs are the Logical-Semantic and the Linking verbs. This finding is consistent with Hinkel's findings that these two are not prevalent in terms of variety, in academic writing.

In order to discuss the types of lexical verbs in more detail, they are listed down below in order of frequency, Activity, Report, Mental/Emotive, Logical-Semantic and Linking verbs. The following results show the most common verbs found in the present corpus, numbered according to frequency.

University of Malaya

Activity verbs

Activity verb is the most frequent verb type used. It has 42% of occurrence in the top 100 verbs. The most common activity verbs registered in the corpus are listed below in decreasing frequency:

1. use	2. show	3. make	4. speak	5. take
6. help	7. produce	8. provide	9. employ	10. get

The list of Activity verbs found in the ESL corpus reflects those from Biber et al.'s findings such as *make, give, take, use, show, produce,* and *provide*. Activity verbs also include verbs that are predominantly used in a cluster or idiomatic forms (*make, give, take*). They can be used in a combination with various prepositions or other words to make up words with different meanings. For example, the verb *make* is most often found in a combination of the word *use*, and the verb *take* is most often found with *place*. It is important to note that the verbs include various forms and combinations with other words because they are highly effective when in use in the combination of two and three-word verbs. Verbs such as *take* are particularly prone to appear in idiomatic expressions or phrasal verbs, e.g: Verb *take* and possible forms. Table 4.1 shows examples of Activity Verb combinations taken from the corpus.

Table 4.1 Verb phrase-based bundles and semantic meaning

Verb	Verb combination	Semantic meaning	Example
take	*take-up*	register	*take up* classes register for classes
	take-place	occur	*takes place* in a classroom occurs in a classroom
	take-on	undertake	massive job to *take on* undertake a massive job

University of Malaya

Analysis of the corpus shows that the writers tend to overuse the phrasal form *make use* which could more appropriately be replaced by the word *utilize*. This example is one of the many that indicate the corpus' population's tendency to utilize phrasal forms which is arguably too simplistic for academic writing. Similar findings were also shared by Pan, Reppen and Biber (2016) in their study of Chinese writers where they are prone to use verb phrase-based bundles when they can be replaced with 'short, more concise and native-like ways'. There are a number of reasons that could have resulted in the use of verb phrase-based bundles such as translation (Halliday, 1989) or the more likely reason as Pan et. al (2016) suggests as lack of academic writing proficiency. Students'genres are more phrasal than the research articles since students depend on using formulaic language (collocations in this study) in developing their arguments more than experts do (Hyland, 2008).

Mental/Emotive Verbs

The second most common verbs found in the corpus is the Mental/Emotive verbs. The most common Mental/Emotive verbs registered in the corpus are listed below in decreasing frequency:

1.find	2.see	3. compare	4. mean	5. observe
6. read	7. understand	8. view	9. identify	10. prove

The highest frequency Mental/Emotive verb is *find*, followed by *see* and *compare*. In raw frequency, *find* occurs over 100 times, which appears to be used to the point of repetition. Although not listed in the top 10, the high usage of the verbs *feel* is another point of interest as they should not be found in an academic prose. One reason that could

University of Malaya

justify use of Emotive verbs such as *feel* is because non-native writers are more exposed to informal, conversational discourse that they find it completely acceptable to use them in their academic writing (Hinkel, 2002, cited in Hinkel, 2004).

Most often, mental verbs are not used as frequently compared to report and activity verbs in academic writing and they represent approximately 0.42% of word count of a large English language corpus (Biber at al., as cited in Hinkel, 2004). It is interesting to note that, contrarily, the corpus produces a high frequency of Mental/Emotive verbs which makes it the second most frequent verb category. A number of reasons could contribute to the high frequency of Mental/Emotive verbs. This study proposes that the reasons could be twofold; the corpus' field of study and the background of the corpus' population.

Considering the population, it has been suggested that non-native writers tend to employ more Mental/Emotive verbs than their native counterparts (Johnson, 1989, as cited in Hinkel, 2004). Mental verbs are often considered as markers for subjective texts, as they are often used to indicate elements of tentativeness and uncertainty (Quirk et al., 1985, cited in Hinkel, 2004). This observation is particularly parallel to the findings made by Hinkel (1997) where she noted that non-native writers tend to be 'vague' and 'ambiguous' compared to the native writers. In addition to influence of informal conversation discourse as mentioned earlier, non-native writers are noted to employ more Mental/Emotive verbs in order to project hesitancy and tentativeness in their claim

The field of study could justifiably be the reason for the comparatively high number of Mental/Emotive verbs. Discussion in the field of Humanities is not based solely on facts but more of the writers' interpretation of the findings that are made sensible via arguments and persuasion. To achieve a balance between over-claiming and weak statements, they have to rely largely on Mental/Emotive verbs.

University of Malaya

Report Verbs

The most common report verbs registered in the corpus are listed below in decreasing frequency:

1.reveal	2.suggest	3. advise	4. discuss	5. report
6. support	7. say	8. write	9. mention	10. state

In this corpus, the most common Report verbs are employed to introduce indirect and reported statements. Indirect statements are given in the form of passive voice *(It is revealed…).* It was also found that reporting verbs are often followed by *that* complementiser. Instances where report verbs are employed with *that* complemetisers are shown in the table below.

Table 4.2 Examples of Report Verbs with *that* Complementiser

Report verb	***that* complementiser**
reveal	the findings *revealed that* repressive humour..
demonstrate	the results *demonstrated that* the speakers' belief
believe	she *believes that* the findings..
say	four of them *stated that* they would

The examples in Table 4.2 exhibit use of report verbs with *that* complemetiser. Two examples are shown with inanimate subjects *(the findings, the results)* while another two examples showcase animate subjects *(she, four of them).* It is assumed that the use of inanimate subjects with active verbs can seem ungrammatical although the notion is rebuked by Hinkle (2004). That is one of the reasons why a passive construct is more popular in academic prose. Further discussion on passive constructs and *that*-construction

University of Malaya

clauses and subject animacy will be discussed later in the chapter under subheading Animate Vs Inanimate subjects and (VP) + that clause fragment.

Logical Semantic Verbs

The most common Logical Semantic verbs registered in the corpus are listed below in decreasing frequency:

1.need (to)	2. represent	3. indicate	4. relate	5. illustrate
6. involve	7. lead (to)	8. increase		

As with Activity verbs and select verbs in the previous categories, verbs in this category more often appear in chunks or collocates. Two out of the eight examples are employed with prepositions. Examples of usage of Logical Semantic verbs are shown below;

Table 4.3 Examples of Logical Semantic Verbs

Logical Semantic Verbs	Example of usage
represents	The firm has cautioned that charity discourse ***represents*** an ideological movement designed to consolidate the power of transnational corporations.
indicate	The high regularity of imperative and hint occurrences ***indicate*** their appropriacy in the advice texts because society would rationally use commonly accepted and suitable expressions in daily interaction rather than unusual structures (Terkourafi, 2005).

University of Malaya

Logical Semantic verbs allow writers to illustrate cause and/or effect by denoting the change in the state of affairs (Hinkel, 2004). In other words, writers employ Linking Verbs to describe the construction of knowledge pertaining to one's research. The patterns of lexical verbs are explored in the later part of this chapter.

Linking Verbs

The least used lexical verbs are the Linking verbs (4 %). It has to be noted that *be* copula is excluded in this study. The most common Linking verbs registered in the corpus are listed below in decreasing frequency:

1.appear	2. seem	3..become	4. portray	5. remain

As shown in the list, the variety of linking verbs found in the corpus is very limited. Linking verbs are not very pervasive in English texts, although their number of occurrences is more dominant in academic texts (Biber et al., 1999). Linking verbs serve as syntactic links that connect subject and subject complements on either side of the verbs (Hinkel, 2004).

The most common Linking verbs are the *be* forms which occur over 20 times more than the other linking verbs (Biber et al., 1999). The list does not include the *be* verbs as they more often mark 'stative constructions' rather than mental processes and actions that take place in discussion sections. Linking verbs are commonly followed by adjectives or noun phrases, the former being more common that the latter. However, Hinkel (2004) voices concerns of overusing Linking verb + Adjective structure, as they may present the text as simplistic and too descriptive. The simple structures of Linking verbs can be

University of Malaya

avoided if the informative phrase follows the subject while the adjective is attached before the subject.

Table 4.4 Concordance Lines with Linking verbs

seem	among the many studies in various contexts, there *seem* to be other others..
	As explained earlier on, the data *seemed* to show that Cleo had…
appear	Items in these areas also *appear* to be related to the notion of a standard…
	contribution generates laughter from the team members and *appears* to be an effective strategy..

Table 4.3 shows examples of Concordance lines with Linking verbs taken from the corpus. They are often used in the present tense and employed as hedges (Hyland, 1998). Linking verbs are associated with the process of reasoning, in which the writers do not want to appear too certain or conclusive. This method of hedging is commonplace in academic writing, and even more so in Humanities prose, where reports are made via persuasion.

4.2 Lexical Verbs and Verb Forms

RQ2: To what extent are the verb forms used in discussion sections in M.A dissertations?

There have been strong suggestions that the studies of verbs should take into consideration both lemma forms as well as verb forms. As proposed by Granger and Paquot (2008, p. 17) “an exclusive focus on lemmas is liable to distort the picture and

University of Malaya

hide some major differences between expert and learner use". A further analysis into the use of verb types and forms can provide an insight into the differences in phraseological patterns favoured by either writer (Granger and Paquot, 2008). With heavy regard to such suggestions, the study also analyses the verbs in the verb form approach. Verbs forms can be analysed in five most basic forms; the VVB, VVN, VVG, VVI, VVZ and VVD.

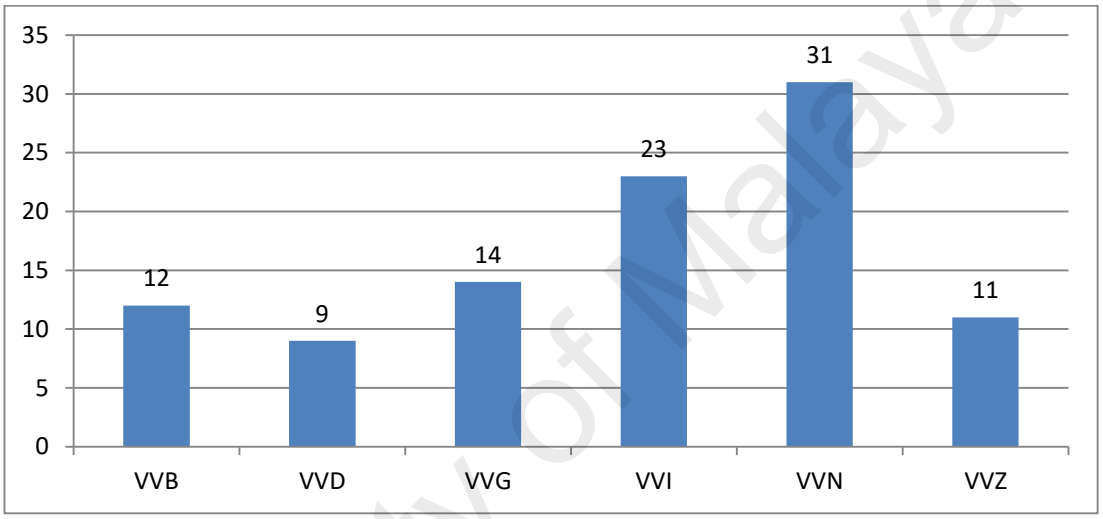


Figure 4.2 Percentage (%) Breakdown of Verb forms

As seen in Figure 4.2, the writers employ verbs in VVN (31%) forms the most, followed by VVI (23%), VVG (14%), VVB (12%), VVZ (11%) and lastly VVD (9%). VVN form indicates the use of past-participle form of the verbs (*shown*, *seen*) which can denote the use of either present perfect or past perfect tense. For reference, abbreviation forms in relation to verbs for CLAWS is also included in this chapter.

Table 4.5 CLAWS Horizontal Output Abbreviation

VVB	base form of lexical verb (except the infinitive)	e.g. SHOW,MAKE
VVD	past tense form of lexical verb	e.g. SHOWED, MADE
VVG	-ing form of lexical verb	e.g.SHOWING, MAKING

University of Malaya

Table continued from page 69

VVI	infinitive of lexical verb	e.g to-SHOW, to-MAKE
VVN	past participle form of lexical verb	e.g. SHOWN, MADE
VVZ	-s form of lexical verb	e.g. SHOWS, MAKES

4.3 Verb forms

The analysis of verb forms allows us to see important aspects of verb use that might otherwise be overlooked if the study only focuses on lemma forms. This is due to distortions that could manifest in two ways; lemmas with similar frequencies will hide overuse or underuse of certain verb forms, and underuse or overuse of certain verb forms affects the frequency of the lemmas (Granger and Paquot, 2008). In any case, an analysis of verb forms presents a more comprehensive measure in the analysis of verbs. Table 4.6 lists the five highest frequency verbs and the breakdown of the verb forms.

Table 4.6 Lemmas and Verb Form Breakdown

Lemmas	**High usage**	**Low Usage**
use	VVI (to use)	VVB (use)
show	VVZ (shows)	VVD (showed)
find	VVN (found), VVI (to find)	VVB (find)
see	VVN (seen)	VVZ (sees)
make	VVI (to make)	VVZ (makes), VVB (make)

As seen in Table 4.6, certain verb forms are more prevalent than the others. The most popular verb forms are VVN followed by VVI for the Top 5 verbs. These verb forms appear to be used with high frequency in the corpus. VVZ, VVB and VVG are revealed to be underused. It is also revealed that certain verb forms are employed almost exclusively that no other forms of verbs are registered for that particular verb lemma. An example of this is the breakdown of lemma *base*, where the only verb form used for it is

University of Malaya

VVN. The overall data on verb forms indicates that the writers' partiality to VVN (perfect) and VVI (infinitive) forms while severely underusing other forms.

4.3.1 Tense

It is suggested that most of the verb phrases are purposely tensed (Biber et al., 1999). For example, in a corpus study of academic genre, the use of the present tense is exponentially prevalent than that of the past tense (Biber et al., 1999). One explanation for this occurrence is that of the functional value of the present tense in stating habitual and often repeated actions. Example 4.1 below illustrates the uses of the present tense for actions.

Example 4.1

1) *The analysis **depicts** that the print media has taken the liberty to inform the reader of the true nature of the report.*

2) *Kierzek (1996) **proposes** that a verb is a word that expresses action, existence and occurrence by combining with a subject to make a sentence.*

In an academic setting, the use of the present tense is viewed as commonplace as the research is not referred to in a specific time-continuum but rather seen as a continuing process that is neither in the past nor in the future. Rhetorical moves in the discussion sections such as reference to previous research, outcomes and explanation as well as making deductions based on the outcomes are relevant examples of the continuous process that develops in a study and thus the use of the present tense is prevalent in the academic writing; especially in the Discussion sections. In the present study, the present

University of Malaya

tense is used in great frequency particularly with these verbs; *show*, *suggest*, *appear*, and *state* mainly for the purpose of stating the findings and outcomes, and referencing.

Concordance lines in present tense

Show	*Since the results **show** that L2 users are more motivated to use the second language when their interlocutors do not switch to English, classroom teachers can teach their students coping strategies for the scenarios in which they do experience switches or how to avoid experiencing switches altogether.*
	*The use of legal borrowing **shows** the limitation of the Malay language.*
Suggest	*These results **suggest** that no one group had an advantage over the other.*
	*However, this finding **suggests** that the quality of sibling play may be related to children's perception of their choice, which did correlate positively with academic and social outcomes.*
Appear	*However, it **appears** that for stress and pausing, the number of errors decreased from the pre-test to the post-test.*
	*On the contrary, questions asked to get information concerning the identity of the prankster or the source **appears** towards the end of the call once the prankster has informed the victim that he/she has participated in a prank call.*

University of Malaya

State	*Meanwhile, Matthews also **states** that subordinating conjunctions join dependent clauses to independent clauses.*
	*Besides that, Wang (2006) also **states** that the Yes/No questions restrict responses and impose more authority.*

4.3.2 Aspect

The present perfect tense signals a continuous process of approaching the subject matter (Wallwork, 2011, as cited in Min, 2013). The perfect aspect can be used in the present and past tenses, but more notably in the present tense. A speaker of English uses a sentence in the present perfect when the information he is giving appropriately exemplifies or explains the topic of discourse (Inoue, 1979). An example is set out in Example 4.2:

Example 4.2

*The term often appears in a particular topic that is being **discussed** and **presented** in the text.*

In Example 4.2, *discuss* is used in the present perfect to refer to the topic which has been discussed and presented earlier in the text. The present perfect is used to show how the problem has been approached from the past until the present day (Wallwork, 2011, as cited in Min, 2013). The present perfect tense indicates something that began in the past (i.e. when research first began in this area) and continues into the present.

The combination of the perfect aspect with a specific tense can create a complex verb form in meaning but one should bear in mind that only a few combinations of tenses and aspects are employed in academic prose. Specifically, according to Biber et al. (1999,

University of Malaya

as cited in Hinkel, 2004) only 8% of all verb phrases in academic discourse are used in the perfect aspect. Various linguists have offered a number of explanations for the use of the perfect tense. Celce-Murcia & Larsen-Freeman (1999, as cited in Min, 2013) explains that the base meaning of perfect is 'prior' and they are often used in context to explain an activity in some other point in time. Examples in the form of concordance lines are shown below:

Example 4.3:

1) *Another significant finding from the interviews is **addressed** by those dealing with international students.*
2) *Furthermore, the essay writing that were **administered** in this study was an argumentative essay which requires...*

In both the examples given, the verb *address* and *administer* have been performed and completed at some point in the past, prior to when the writing of text takes place. In the ESL corpus, the perfect aspect is mostly used with actions that are necessary to be carried out in the process of conducting a study; for example, identifying and addressing problems, administering survey questions and measuring the scales of previous findings. It appears that no distinction between the use of present and past perfect aspect has been made in the examples of their usage.

The quantitative variations of the verb forms indicate the writers' partiality of different phraseological patternings. Clearly, ESL writers studied in the present corpus favour the VVN form, deviating from Biber's (1999) observation as well as Granger and Paquot's (2008) study where the learners' corpus (the ICLE), where the most frequently

University of Malaya

used verb form is the VVI. VVI forms indicate the infinitive form of the lexical verbs (*to see, to show*). In the current learners' corpus, the VVI is the second highest frequency verb form. The third verb form is the VVG (*verb+ing*). The least used form is the VVD. Other than in historical or biographical texts, past tense verbs tend to occur in small numbers across academic fields (Quirk et al., 1985, cited in Hinkel, 2004).

4.3.3 Infinitives and gerunds

The percentage of verb forms in Figure 4.2 shows that VVI forms (infinitives) are more favoured compared to the VVG (forms). This is also true in the study by Granger and Paquot (2008) where it was found that the VVI forms generate more frequency than the VVG forms. Table 4.7 reveal common verbs for the highest frequency of VVI and VVG forms respectively.

Table 4.7 Verbs in Infinitives and Gerunds

VVI (Infinitive)	Frequency	VVG (Gerund)	Frequency
use	20	use	37
construct	18	make	28
produce	9	construct	26

The data shows that there are two verbs *(use, construct*) which are most frequently employed in both the VVI and VVG forms. As noted by Conti (2011), non-native writers especially, find the choice of the infinitive and the gerund as an 'arduous grammar point'. The findings from this study brings forth the old age question about whether to use the infinitive or the gerund for complements. Further analysis is performed through an examination of the concordance line for the verb, *use*.

University of Malaya

Infinitive of use	Gerund of use
*The essay analysis showed that only seven students attempted **to use** more than 10% of the newspaper wordlist in their writing.*	*According to the study, the professionals are not able to practice **using** the language.*
*They appear to have good understanding in parts of speech and are able **to use** the words correctly in elaborating their ideas.*	*The purpose of the advertisement is to enable the consumers to believe on the impact of **using** the products based on the positive language used.*

Previous studies (Conti, 2011; Arseneau & Duffley, 2016) have noted that for certain purposes for the verb, *use,* the infinitive forms and gerund are generally interchangeable.

4.3.4 Passive and active voice

As discussed in the previous chapters, the use of the passive voice is generally seen as the convention in academic writing and much more common in academic prose although it is also noted that many academic writers are also partial to the active voice because it is more direct and concise (Biber et al., 1999). For example, a study of Indonesian learners' theses by Yannuar et al. (2014) shows that the active voice is used more frequently (64%) compared to the passive in the analysis of stance verbs (such as suggest, expect, show, predict and report). However, it must be noted that the rest of the stance verbs (36%) are constructed in the passive voice. In this study, the percentage of usage of both voices is almost equal. The percentage of the passive voice is calculated based on the occurrence of VVN forms while the active voice is identified by the use of

University of Malaya

VVB, VVZ and VVD forms (refer to Figure 4.2) which amount to 31% and 32% respectively.

Biber et al. (1999) notes that both voices can be employed interchangeably, in which case the authors may choose a voice that can fulfil a certain purpose in discourse functions. In other words, both voices are useful and necessary and that academic writers need to equip themselves with the skills and the knowledge to apply both voices appropriately. The passive voice is constructed by placing the verb-to-be in the same tense as the active verb and replacing the active verb with the past participle form. The choice to use the passive voice also contributes to a greater frequency of the VVN form. Biber et al. (1999) remarks that academic writers are more partial to the active voice as they convey a direct and concise message while both voices are still considered useful and necessary in the academic context as they serve different purposes that could reflect the writers' intent. Example 4.4 shows the concordance lines using the passive voice.

Example 4.4

1) *Such variety has not reached a stabilised form of use **accepted** by all the members in the speech community*

2) *The women will also try to look beautiful and gorgeous like what has been **depicted** by the models in the television.*

In both the examples, the passive voice is constructed by placing the verb-to-be in the same tense as the active verb. The active verb is changed into the past participle form. This method is often utilized to switch the subject of the verbs to become the agent of the passive verb. In many cases, the agent is introduced by using the preposition *by.*

University of Malaya

As a rule of thumb, the passive voice is employed when the writer deems that the action is more important than the agent. This situation often occurs particularly in academic prose, reports and processes.

However, the more common reason to use the passive voice is to avoid mentioning the agent altogether. The agents in these instances are often the authors themselves, when, in the spirit of anonymity and conventionality chose the passive construct to avoid any reference to themselves. For these reasons, the use of the passive voice is generally considered as the 'academic discourse convention'. Example 4.5 illustrates instances where the passive voice is employed in order to avoid mentioning the author.

Example 4.5

1) *At the same time, such limitations cannot be **avoided** without incurring certain cost.*
2) *These guidelines are a limitation to the framework as some of the categories in the surface structure taxonomy cannot then be **applied**.*

As mentioned previously, native writers tend to use the active voice more in their constructs despite their conventionalised use in academic writing. The frequent use of the passive voice could also indicate some cultural elements at play. Further discussion regarding cultural impact on academic writing will be discussed in Chapter 5.

4.4 Associated patterns

RQ 3: What are the associated patterns of lexical verbs used in discussion sections in M.A dissertations?

University of Malaya

Further investigation of the verbs found in the corpus leads to the examination of the phraseological patternings that shape the verb clauses. As mentioned previously, the verbs tend to appear in phraseological forms. The structural patterns (based on Biber's classification) that are observed in the present study are as shown in Table 4.8:

Table 4.8 Verb-based Lexical Bundles in Raw Frequency

Lexical bundle	Raw frequency	Example of Concordance line
Passive verb + PP fragment	703	*The theoretical framework was used in..*
(VP) + that clause fragment	235	*The findings shows that...*

Table 4.8 shows verb-based lexical bundles in raw frequency. Note that for a relatively small corpus such as this, a raw frequency is often used (e.g., Altenberg, 1998; De Cock, 1998, cited in Chen and Baker, 2010). Passive verb + PP fragment is more commonly used compared to the (VP) + that clause fragment.

While it is worthwhile to see that the raw frequency counts of each verb use and to compare the numbers in the three groups, the numbers themselves are preliminary for analyzing the patterns of ESL usage of the verb phrase without actual samples. Therefore, the tables should not be over-interpreted, but used as a complementary data, when scrutinizing the phenomena case by case with the actual student samples because only then, the numbers counted for each use would make sense.

4.4.1 Passive verb + PP fragment

Table 4.9 shows recurring pattern involving Passive verb + PP fragment.

Table 4.9 Passive Verb + PP Fragment in Raw Frequency

Cluster	Frequency	Example of Concordance line

University of Malaya

Table continued from page 79

Passive verb + in	260	*These factors were **cited in**....*
Passive verb + by	226	*Humor is often **accompanied by**...*
Passive verb + to	128	*It may have been **aimed to**....*
Passive verb + on	89	*The inference was **based on**...*

As shown in Table 4.9, the verb-based pattern Passive verb + PP fragment is employed with a comparatively high frequency. The most common combination of Passive verbs are with Preposition *in* followed by Preposition *by, to* and *on*. As noted earlier, passive constructs (VVN forms) are conventionalised in academic writing, as evidenced in the present corpus. Passive voice is commonly utilized in academic writing, to an extent that it can be seen that the academic discorse conventions are promoting such rhetorics (Hinkel, 2004b; Poole, 1991; Swales, 1990). Passive voice is used in academic writing as a platform that serves multiple textual functions in accordance to the author. One of these is to project an academic indirectness, detachment, and objectivity that is requisite in English-language academic tradition, and particularly so in natural sciences and engineering (Hinkel, 1997, 1999, Johns, 1997). Based on a large number of corpus analyses of academic prose, the passive voice is ubiquitous and remains a prevalent feature of academic text in various disciplines (Biber, 1988; Hyland, 1996; Johns, 1997; Swales, 1990). Example 4.6 shows the use of the Passive verb + PP fragment.

Example 4.6

1) *The results in this study proved that code-switching is one of the appropriate strategies to teach students with non-native English speaking background in the school where the research was **conducted.***
2) *Regarding the students interviews in this study, the findings are mostly in accordance to the theory **proposed** by Crystal.*

University of Malaya

3) *The first research question was formed to find out the types of questions that are **constructed** in the Gotcha prank calls.*

4) *In the researcher's opinion, this scenario could be **associated** with the students' self-perception of their own abilities.*

It is important to note that the majority of the Passive verb + PP fragment combination is specially utilized to omit the agents or the subjects, with the exception of *Passive verb + by* fragment. The concordance lines 1, 3 and 4 in Example 4.6 have chosen to omit the agent who are likely to refer to the researchers themselves which indicate detachment. Contrarily, concordance line 2 was clearly constructed to mention the agent.

The combination of *Passive verb + in* is most commonly used in the form of, *cited in,* by way of avoiding to mention the researcher. A similar approach is observed by using the *Passive verb + to* and the *Passive verb + on* where the most common combination is *compared to* and *based on* respectively. These patterns are commonly found in the academic genre; especially, the Humanities discipline where the academic prose tends to be persuasive and argumentative in nature, thus requires authors to compare and contrast their findings with previous works (Maroko, 2013).

Concordance line 2 in Example 4.4 presents a different approach in using the Passive verb + PP fragment. In the example, the combination of Passive verb + by allows the writer to specify the agent by placing them as the objects instead of the subjects. It is interesting to note that the use of prepositions, including the *by* preposition in passive constructs indicates a somewhat contrary approach towards the agents of the phrase, but somehow to achieve the same goal; detachment and academic indirectness. On the one hand, in short passives with the preposition *in* such as *shown in*, the agents are not specified at all, though, by using the preposition *by* in a long passive, the authors are able

University of Malaya

to specify an agent although it reduces the significance of the agents. In most cases of a research paper, the agent could refer to the researchers who are working the study, thus eliminating the need to mention them.

4.4.1.1 Short passive and long passive

Short passives generally recorded more use, compared to the long passive (Biber et al., 1999) which can be employed in two forms; with two object prepositional verbs, corresponding to an indirect object of a ditransitive verb or prepositional object of one object prepositional verb. Example 4.7 shows a short passive with one prepositional object:

Example 4.7

1) *The verbs are usually **classified** as lexical, copula and modal verbs.*
2) *They are **given** as an example of collocations.*
3) *The suggestion was **received** with mixed response.*

In example 4.7, short passives are used with one prepositional object (example of collocation, mixed response). In the first example, *lexical, copula* and *modal verbs* are referred to as one prepositional object.

Ditransitive verbs are verbs that correspond with two objects, a direct and an indirect one. In the case of passive construction, it is possible to be formed by placing both the direct and indirect objects in the subject position. To make the explanation clearer, this study adopts Huddleston et al.'s (2002) classification of the objects which are first passive and second passive. The first passive is identified as the indirect object which is placed in the position of the subject while the second passive is identified when the direct object is placed in the position of subject. With the ditransitive verb structure,

University of Malaya

Huddleston at al. (2002) propose that the first passive is more prevalent than the other. Note that not all the verbs can be used as ditransitive verbs. The most common ditransitive verb found in the corpus is *give*. Example 4.8 exhibits concordance lines with ditransitive verbs.

Example 4.8

1) *These repeating themes could have **given** Thelma the practice, experience and ability to identify the forms that work best with specific problems and questioners where they probably responded more positively to the advice when imperatives and hints are used as Thelma manages to convey sincerity, friendliness and formality simultaneously.*
2) *Mainly because very little attention have been **given** to the text producers' choice of the words and the generic structure of the texts.*

There could be a number of reasons why the writers have chosen to place the indirect object in such a away but it is generally seen that the formation of the first passive is seen to serve as an emphasis on the indirect object. Such emphasis could make readers more attune to the arguments the writers are making. In the example, it is seen that the writers choose to place the seemingly unimportant detail (*These repeating themes, very little attention)* as the subject but those details are actually the main point of the sentence.

Long passives (with a *by*-phrase) are used with less frequency than the Short passives, as they are much less common in academic constructs but evidently appear frequently in the present corpus. Long passives can easily be reconstructed with an active clause, although their use is much more motivated by these three principles: Information-Flow Principle, End-Weight Principle and to place emphasis on an element of the clause

University of Malaya

(Biber, 2006 p. 179). These principles assumed by Hinkel (2004) to serve as a cohesive chain where in context in which the agent is unknown or is rendered unimportant, the *by-phrase* can be omitted (in which case referred to as the Short Passive). Long Passives (as cohesive devices) taken from the corpus are shown in Example 4.9.

Example 4.9

1) *.....vocabulary acquisition than those learners who are exposed to pre-modified input only is confirmed by* ***the findings of the study****.* ***All input*** *conditions showed reasonable level of receptive vocabulary...*

2) *...features and the generic structure of the texts are significant resemblance of the discourses drawn by the* ***text producers*** *in producing the texts.* ***They*** *create the different schemas which are drawn not..*

Example 4.9 presents the use of long passives as cohesive devices. Example 1, instead of placing the *findings of the study* as the subject, it is moved to the back as the object. In the next sentence, it is employed as the subject (*all input)* to ensure cohesion between the two sentences. A similar strategy is employed for Example 2 where the *text producer* is placed in object position and the next sentence starts with the pronoun *they* which refers back to the text producers.

A majority of the passive constructs are idiomatic, in this case, with the prepositions *in* and *by* which can be problematic for L2 learners, as noted by Atkinson (1991) and Owen (1993, as cited in Hinkel, 2004). Contrarily, the advanced ESL writers in the present study appear to utilize both constructions in a similar frequency.

University of Malaya

4.4.2 (VP) + that clause fragment

Among the patterns of verb forms, the (VP) + that clause fragment form has been a subject of interest for a large number of studies, in particular as part of Verb-based lexical bundles (Bal, 2010; Chen, 2010; Bungor 2016). It occurs predominantly in reporting clauses as noted by Charles (2006). The same observation was made by Hinkel (2004) in all her examples of Report verbs followed by the complementiser *that*.

Previous studies on (VP) + that clause fragment have always focused on the writers' stance and attitudes in reporting their claims (Hinkel, 2004; Charles; 2006; Liu, 2014). Therefore, this study also analyses the *that construction* from the reporting perspective. The table below lists the Report verbs that are most frequently used with *tha*t.

Table 4.10 Common Report Verbs

Report verb	Raw frequency	Report verb	Raw frequency
show	65	state	13
find	31	observe	9
reveal	30	prove	7
suggest	23		

Table 4.8 shows the most common report verbs used with *that* construction found in the corpus. Note that the ***focus*** of this paper is the discussion section; hence, the smaller number and less variation of reporting verbs. The highest frequency report verb paired with *that* construction is *show (n=65),* followed by *find, reveal and suggest.* The high frequency of *show* in the corpus echoes concerns from previous studies that non-native writers are prone to repetition notably in verb use. Example 4.10 highlights the concordance line of *show* whose use is notably repetitive in the corpus.

University of Malaya

Example 4.10

1) *This **shows** that the minial use of emoticons in Twitter is in line with...*

2) *Their results **showed** that most of their subjects made errors in terms of omission...*

3) *The present data **shows** that there is a difference on the length of tweets by both writers*

Examples 4.10 of concordance lines 1, 2 and 3 all reveal that *show* is commonly used to introduce data and findings i.e an inanimate subject. It also appears that the writers' favour certain verbs to be paired with inanimate objects and vice versa. This phenomenon is explained in the later part of the chapter.

Hyland in his studies and publications (1999b, 2000b, 2002a, 2005, 2008) has been very keen to examine the writers' stance and its relation to social interaction in academic writing, with specific focus on the analysis of reporting verbs. The examination of the writers' stance is strongly associated with report verbs, as the verbs have to be chosen carefully in order to convey the writers' rhetorical purposes. In the **Discussion section**, the writers are compelled to perform an evaluation of their own findings. In this case, an analysis of the Epistemic stance device (Biber, 2006, as cited in Ağçam, 2015) is most applicable, in particular the *verb+ that* clause. It was argued by Hyland (2004, as cited in Ağçam, 2015) that epistemic devices are salient elements in academic writing as they enable the writers to perform a self-evaluation of their own findings and statements and that readers are able to read the writers' uncertainty of deference in subtext. Table 4.11 shows categorisation of the Report verbs into classification of Epistemic verbs.

University of Malaya

Table 4.11 Categorization of Report Verbs as Epistemic Verbs

Certainty	Likelihood
show, find, reveal, observe, state, prove	*Suggest*

In Table 4.11, the report verbs are categorised into 2 types of Epistemic verbs, one that conveys certainty and the other one that hints likelihood. 6 out of 7 verbs are categorised as *certainty* while the remaining *suggest* are placed in the *likelihood* category. It is safe to conclude that the population of the corpus is more partial to *certainty verbs*, which indicate their positive and confident attitude towards their own findings and statements. Note that verbs *feel* and *believe* are also recorded in the form of *that construction* although they are not present in Biber's (2004) register of Epistemic verbs. As discussed previously, ESL writers in this corpus are noted to use colloquialisms in the academic prose. Nevertheless, it could be an interesting endeavour to study deeper into this peculiarity. A similar finding was also observed in the study of Turkish Academic Corpus of English (TACE) by Ağçam (2015) in terms of a higher percentage of *Certainty* verbs and the presence of peculiar verbs such as *believe*. Example 4.11 shows concordance lines with verb *feel* and *believe*.

Example 4.11

1) *The writers **believe** that the reason why errors were made in the subject verb agreement...*
2) *The researcher **feels** that apart from the methodologies that can be used...*

Further analysis into the concordance lines shows that the writers tend to use *Likelihood* verbs to refer to their own statements, in the manner of hedging while

University of Malaya

Certainty verbs, such as *find* and *reveal* are almost exclusively used to introduce hard facts such as findings and data. In humanities, the writers are obliged to make statements that could relate their findings to previous studies. This is especially true as remarked by Hyland (2002, p. 11) in determining the distinction of citation and reporting culture of hard and soft sciences where the soft science is required to "engage in more recursive patterns of investigation which involve more diverse and less predictable and abstract subjects than those typically found in the sciences." It has also been noted that non-native writers are more inclined to avoid strong verbs as to deflect from making 'strong statements'. In this situation, it appears that both elements have come to play through the choice of the verbs *feel* and *believe*.

4.4.2.1 Tense of Reporting

Cluster patterns identification using WordSmith Tools (Scott, 1999) reveals that *that* is commonly paired with finite verbs which could indicate choice of tenses. Choice of tense signals varying attitudes and distance the writer projects based on their findings and statements (Swales, 1990). Tense alternation is also seen as a 'strategic communicative device' that allows writers to communicate their arguments to the readers (Sakita 2002, as cited in Chen, 2009). Subsequently, the choice of reporting tense may differ across chapters. In a study by Chen (2009) in a review of tense use in Literature review, it was found that the majority (70.6%) of the reporting verbs are in simple present forms, 8.9% in simple past and the other tenses account for 4.7%.

Table 4.10 Tenses of *that construction* Clause

Tense	Raw Frequency	Percentage (%)
Present	169	57
Past	128	43

University of Malaya

Table 4.10 shows a choice of tense for *that construction* in the Discussion sections of M.A. dissertations. The simple present carries 57% while the simple past carries 43% respectively. Although the simple present is employed more frequently than the simple past, the difference is not very significant. Example 4.12 shows concordance lines with the simple present and the simple past.

Example 4.12

*1) The second theme **shows that** perceptions of the local variety...*

*2) Most participants **believed that** some people have a natural tendency to..*

The general convention of tense usage in citations and reporting may vary according to disciplines. Humanities favours the *simple present* and the Sciences tend to employ the *past tense* (Maroko, 2013). However, it is noted that academic construct requires the use of the present tense in the discussion of published theories and findings. Such published work is generally considered to be established knowledge and the use of the *present simple* reflects this (Wallwork, 2011). Example 4.13 identifies the use of the simple present tense.

Example 4.13

1) Therefore, this clearly ***shows that*** Malaysiakini provides a much more….
2) Swain (1995) ***argues that*** the processes involved in comprehending a message…

The first concordance line exemplifies the use of the *simple present* across the corpus, as a conventionalized tense to report present data in their research. The second

University of Malaya

example refers to the occurrence during an interview carried out by the author, in which case the *simple past* should be used. As remarked by Wallwork (2011, as cited in Min, 2013) "the past simple is required because the actions you describe took place in the past (i.e. before you started to write your paper). The *past simple* also helps to distinguish what you did from what others have done". The *simple present* is employed following the general convention of academic writing where it is used prominently across the corpus. As identified in the second line in Example 4.12, the writer is referring to a published theory in the field of study, in which case the *present tense* is employed.

However, a high percentage of *past tense* use may be a point of concern. It appears that they also employ the *past tense* forms in a rather high frequency. It shows that the convention is not fully adhered to in this corpus. This finding supports Hinkel's (2004) claim that non-native writers are partial to *past tense* narration when explaining and supporting their arguments. Example 4.13 exemplifies the use of both tenses in the discussion of their findings.

Example 4.14

1) *In this study, the researcher **was interested** in investigating the reasons and functions of code switching in a secondary school in Kuala Lumpur, Malaysia. The data **was gathered** from a questionnaire, observations and interviews that elicit reasons and functions of code switching. The findings **proved** that teachers and students used code switching in ESL classroom.*
2) *The premier objective of this study is to distinguish the different portrayal of two*

University of Malaya

different social actors in news reports of wife abuse, an issue of domestic violence. The two social actors, who are the victims and the perpetrators, ***are being portrayed*** *differently by the text producers, who* ***are reporting*** *on the cases.*

In Example 1, the writer chose to employ the *simple past* in the narration of her discussion. This writer's choice may be explained that she follows the timeline, in which, the findings and the arguments are made during the process of data analysis, and not during the write up of the paper. Upon further examination, it was found that there has been an excessive generalization among renowned academic scholars with regards to the appropriate tense choice in academic writing. This view has been voiced by Min (2013) where she lists several generalizations regarding the use of the *simple past* in an academic context. On the one hand, Wallwork (2011) suggests that the *simple past* is used in the context of describing one's findings. Other scholars such as Wray (2009) recommends that the *simple pas*t be used to refer to previous findings and statements in order to clearly situate the present study within the context of the established knowledge. Such a broad generalization is applicable across discourses and disciplines, and is often without specific examples in terms of discourse and verbs have left non-native writers to decide on sentence level tense alternation or paragraph level tense alternation (Min, 2013). In other words, writers may apply whichever tense they feel appropriate for a particular sentence or paragraph, often disregarding the consideration for discourse or context. Example 4.15 highlights the occurrence of paragraph level tense alternation.

Example 4.15

1) ***Fairclough (1992b) claimed that****, texts may be sincere as well as manipulative. The disabled people have been positioned in limited ways which echoes Croteau & Hoynes (2000:166) claim that the media engage in practices*

University of Malaya

that 'define reality'. Social construction is a process of actively creating meaning; media images do not simply reflect the world, they represent it (Croteau & Hoynes, 2000) (cf. Section 1.7). Certain aspects of reality are highlighted and others neglected in the process of text creation. Hence, representations of the disabled can be incomplete and narrow. Sociologists, social theorists and political economists building on the earlier works of such as ***Marx, Weber, Habermas and Marcuse have cautioned*** *on the increasing penetration of the social by the economic in societies (Banarjee, 2006). To sum up,* ***Barnes (1997) considers*** *the social model of disability being linked to two traditions – social construction (cf. Section 1.7) and the 'social creation' of industrial capitalism (Barnes, 1997:5) (cf. Section 2.5.2).* ***Corbett's (1996) conclusion that the construction of social reality of the disabled community is a social conditioning*** *phenomenon and Foucault's idea that disability is a form of subjection (cf. Section 4.4.2) are in line with Fairclough's claim that language is a socially conditioned process, conditioned by other non-linguistic parts of society.*

2) *Appeals to ethos were evidently* ***employed*** *at Introductory Stage though the scammer still consistently* ***maintained*** *his credible persona throughout the correspondence, especially by using language which connotes religious orientation throughout the correspondence to maintain his religious persona, as well as other appeals to logos to proof that he is trustworthy and reliable. Essentially, this is expected as the moment the scammer* ***loses*** *the targets' trust, his entire scheme would end. His display credibility the form of trustworthiness, high reputation and morally-upright attributes in the Introductory Stage also*

University of Malaya

serves *the purpose of attracting the target (in addition to the other attractive elements that he would have included in his fake online dating site profile).*

In example 4.15, it is observable that the writer alternates the tenses with little consideration of the overarching discourse grammatical rules. In the first example, tense alternation occurs among verbs referring to previous studies. As mentioned earlier, it is generally proposed that the *simple present* is used to refer to previous literature as it implies the notion of established knowledge. However, in the first citation of the paragraph, it appears that the writer is using the *past tense* form while the following citations follow the discourse convention. The second example also exhibits a slight mix-up of tenses. While the writer seems to apply past tense to refer to his own findings, one verb *lose* is employed in present tense.

The perplexing ways scholars have outlined tense usage appears to cause confusion among the writers within this context. One needs to bear in mind that the broad generalization of tense usage in academic genre need to conform to the discipline in which they are employed. Otherwise, tense mix-up as highlighted in the examples could occur. Other factors that could contribute to problematic tense use is *lack of intuition on the writers' part*, primarily with writers from tenseless L1 background (Hinkel, 2004) such as the population in this corpus. In her study of speakers of tenseless languages such as Chinese and Japanese, she identified that writers from such background find that the *simple past* is the easier tense of the English language to acquire, thus promoting their use in writing.

University of Malaya

4.4.2.2 Animate vs Inanimate subjects

The *v_that* patttern is also indicative of reporting and referencing culture in the corpus population. In referencing, Hinkel (2004) argues that the animacy of the subject (human or non-human) does not in any way affect the grammaticality of the reporting clause. However, for many writers, especially from the L2 background, citing an inanimate subject may be considered ungrammatical. The table below show breakdown of subjects of Report verbs into animate and inanimate subjects.

Table 4.12 Subject of V_that construction by frequency

Verb	Raw frequency	Animate	Inanimate
suggest	39	7	32
show	60	1	59
argue	9	8	1

Table 4.12 illustrates the choice of subject for select *v_that* phrases. The concordance lines from the corpus suggest writers' inclination to pair verbs according to the *animacy* of the subjects. The corpus employ V (show) *that* with largely inanimate subjects, particularly referring to findings and tables. *Argue* are dominantly paired with animate subjects and *suggest* are used with both types of subjects as illustrated in Example 4.16.

Example 4.16:

1) *Table 4.1* ***shows that*** *there are certain..*
2) *Although Child (1992)* ***argued that*** *it was not easy for*
3) *The author* ***suggested that*** *formality at..*

University of Malaya

*4) The data **suggests that** the person..*

This result proposes that the writers do not have issues regarding the *animacy* of the subjects despite some researchers' concerns, although they do discriminate the choice of verbs to use with either animate or inanimate subjects.

A further analysis of *subject animacy* and reporting verbs reveals a peculiar pattern of the ***V_that*** clause tense. The patterns show a link that indicate that the *present tense* is largely employed when reporting the present data (inanimate) while the past tense is used when citing previous works. The simple present in the citations could reflect the current state of knowledge and the present implications of research findings that form part of the character of knowledge construction in the humanities and social sciences. However, it is noted that academic construct requires the use of the *present tense* in discussion of published theories and findings. Such published work is generally considered to be established knowledge and the use of the present simple reflects this (Wallwork, 2011).

4.5 Summary

The chapter has presented the findings from the corpus. The findings are divided into three subchapters which are *verb types, verb forms and verb patterns*. Identification of verb types is based on Hinkel's (2004) classification of academic verbs. The present study examined and analysed *the use of lexical verbs* in the Discussion section of M.A dissertations. The findings are parallel to previous research where *Activity verbs* are the most frequently used verbs. However, it differs greatly in high frequency of Mental/Emotive verbs, followed by Report Verbs, Logical-Semantic verbs and Linking verbs. The analysis of the verb forms that the ESL writers in the corpus favoured showed that the VVN forms were the most favoured, followed by VVI, VVG, VVB, VVZ and

University of Malaya

lastly VVD. The results indicate the writers' tendency to employ the passive voice in their writing, which is generally avoided by ESL and other non-native writers in other studies despite its conventionalised use in the academic discourse. The verbs associated patterns were carried out through the analysis of Passive+PP and V-that clauses. It was found that the Passive voice is frequently followed by PP *in* and *by*. It indicates the writers' differing methods of reducing the visibility of the subject and thus reducing its significance. In both instances of PP *in* and *by*, it is clearly presented that the writers employ this method to comply to conventions of academic writing which is objectivity and detachment on the writers' part. The analysis of *V_that* pattern showed writers' referencing and reporting culture of the corpus population. The corpus revealed a tendency to discriminate verb types according to the animacy of the subject. Common *V_that* patterns with verb SHOW, ARGUE and SUGGEST revealed that SHOW is almost exclusively used with inanimate subjects, ARGUE is used with human subjects and SUGGEST is used with either types. A more thorough discussion on significant findings from the analysis is provided in Chapter 5 along with the impact of the findings and the pedagogical implications of the present study.

University of Malaya

CHAPTER 5: DISCUSSION, IMPLICATIONS AND CONCLUSION

5.1 Introduction

This study focuses on examination of lexical verbs and lemmas as well as patterns highly associated with verbs. In the previous chapter, lexical verbs are classified into five types of verbs, verb forms are examined and common verb patterns are identified. Lexical verbs, while highly salient in any forms of writing, especially in academic writing has received considerably less focus. Lexical verbs are particularly relevant in use during discussion where arguments and persuasion are organized through manipulation of verbs. The corpus, which consists of M. A dissertation, follows more subjective approach to making conclusion. Unlike the hard sciences, verbs are employed as tools to help readers to understand their arguments and to an extent, encourage them to agree with the authors. Therefore, it is essential that use of verbs is examined closely to reveal how they make an impact to the discussion in corpus.

5.2 Types of verbs

The Top 100 verbs identified in the corpus are categorized into 5 types of verbs- Activity, Report, Mental-Emotive, Logico-Semantic and Linking verb. The order of the most common verbs by category is Activity, Mental-Emotive, Report, Logico-Semantic and Linking Verbs.

It is concurred that Activity verbs are generally more prevalent in many genre of writing, this includes academic prose. Activity verbs are usually used to explain various process and procedures implemented during a study thus making them more salient particularly in the methodology chapter (*use, employ, produce*). In discussion section, these verbs are also quiet prominent. The highest frequency Activity verb is *use*, which is appears to be overused in the corpus. This is because the discussion follows certain

University of Malaya

pattern in which they narrate the process occurring in the methodology section in order to describe their findings. It is also worth noting that the subject of study in most the dissertation are related to people and their language use, and these factor contribute to high usage if use (..*use the language to*..) Interestingly, the verb *employ*, which are essentially synonyms, are also highly common. Other synonyms of *use* such as *apply* and *utilize* did not make the list. Second most common Activity verb is *show*, which are used to refer to the findings and data from their own study. In Discussion section, it was found that simple Activity verbs such *as make, take* and *get* are not often used on their own but in collocational forms such as *make sure, take place* etc. Language scholars have cautioned against using this form of verbs, as they signify colloquialism and inaccuracy in reporting. Swales (2004, as cited in Granger and Paquot, 2008) argues that no matter how complex the data is, the use of informal English may render it too simplistic. This paper intends to question the motive behind using these phrases as they can be easily replaced by more succinct and precise verbs (*make sure* to *ensure*). However, it is not possible without interviewing the corpus population.

Mental-Emotive verbs occurs second in frequency after Activity and followed by Report verbs. This particular finding in itself is quite alarming as Emotive-Mental verbs are noted to be sparsely used in academic construct (Biber, 2009). Academic piece should normally incite confidence in their findings and to register large number of such verbs can be a cause for concern. On the other hand, it appears that this phenomenon is rather a normal occurrence after taking into account these two variables; field of study and the corpus population.

Mental-Emotive verbs are noted as subjective markers and they can be manipulated to affect tentativeness and uncertainty. Aptly, the Humanities is the study of human and their idiosyncrasies, and in this the results and findings should not be taken as

University of Malaya

definite. It is crucial that the arguments and conclusions are not presented too strongly as the results could easily be interpreted and manipulated in a different direction. Researchers could make arguments and suggestions for their conclusions, though they are still subject to interpretation and persuasion.

The corpus population consists of Malaysian ESL writers; whom by all account and purpose are considered non-native writers. Non-native writers are found to have the tendency to be vague and ambiguous in their writing. The reason for this could be cultural; the Malaysian cultures frown upon appearing too confident and assertive. It could be that they worry that they might repel the readers if they appear too strong or they want to communicate their empathy. In this way, the mental-emotive verbs are great tools to make their discussion appear less assertive and more empathetic. The cause for concern for language teachers and instructors alike are the use of conversational verbs in academic writing. Verbs such as *feel* and *believe* are known as conversational verbs because they are almost exclusively used in conversations, and rarely in formal, academic prose. Perhaps it is one of the method employ to gain readers' empathy, or they are more exposed to informal language use that they find it acceptable employ these verbs in their academic writing. This paper suggests than the reason behind the use of conversational verb is the former rather than latter, considering the advanced ESL population of this corpus.

Logical-semantic verbs are employed to show relationship between actions and event or cause and effect (Halliday, 1994, as cited in Hinkel, 2004). As shown in the findings, this type of verbs often occurs in collocational expressions. This result is also shared by Hinkel (2004) in her list of Logical-Sematic verbs.

University of Malaya

Verb	Example of Expression
Cause	*Cause concern, cause problems, cause incovenience*
Combine	*Combine to do xxx, combined effort/action, combined with xxx*
Develop	*Develop an idea, develop an illness, develop a relationship*

(Taken from Hinkel, 2004)

In view of this finding, Nation (2001) and Wray (2002) suggest learning of collocational expression in context. Contextually learning for verbs such as these is crucial as they are hardly encountered in other genres, to the extent that some expressions are only specifically found in academic texts.

Linking verbs are more prevalent in academic text than other written genres. However, the most popular form of linking verb is the *be* form. Due to this, there is little variety of linking verbs, although their use is dominant. The most common linking verbs are *appear* and *seem*, which are also widely used in this corpus. The linking verbs employed in this corpus appear to serve two purposes, as both linking verbs and hedges. Hedges are widely used in academic writing, even more so in the Humanities. Some scholars encouraged the use of hedges in order to minimize the overly strong suggestion and overgeneralization.

5.3 Usage of verb forms

Overusing and underusing certain verbs or verb forms are also an area of interest among language scholars, especially in non-native writing. As pointed out earlier, an observation that is shared by many scholars is that non-native writers often display limited vocabulary range to the point of repetition in compensation of their limited vocabulary

University of Malaya

and thinking in L2 whereas their Native counterpart are more prone to variation (Jin, 2008). Based on her own study, Hinkel (2004) suggests that ESL writers' texts tend to be built on restricted vocabulary range, to the point where their writing appears repetitive and constrained. While it is true that some verbs tend to repeat itself in academic discourse, such as relatively simple verbs such as make, do and look, and some verbs are identified as the common academic verbs (e.g *identify, apply, investigate*), there are limits as to when it can appear overuse or underuse in an academic prose. Another study on lexical verbs, more specifically reporting verbs at a local university by Manan and Mohd Noor (2014) reveals by that non-native writers are prone to repeat certain category of verbs compared to other categories of reporting verb in their theses.

In this study, the most oft repeated verbs are from the Activity verb category*; use* and *show*. The higher frequency of use for these verbs indicate lack of variety of other synonyms which can easily replace these verbs. Table 5.2 exhibit synonyms for verbs *use* and *show*;

Table 5.1 High Frequency verbs and synonyms

Verb	**Synonyms**
Use	Appropriate, employ, adapt, utilize
Show	demonstrate, display, exhibit, present, reveal, expose

(Taken from thesaurus.com)

Further reading into lexical choice of non-native writers reveals contrasting findings in previous literature. While some studies argue that the more specific words the writer uses, the more proficient they are, contrastive view states that as more time spent of in studying a language, the words used become less specific. To clear the contentious arguments, Crossley et al. (2011) counted the number of different words, which resulted

University of Malaya

in the conclusion that advanced learners use more specific words and different ones, where beginners use general words and have limited lexical variety.

5.4 Analysis of Verb forms

Previous studies have suggested that writers from non-native background of various proficiency levels displayed the tendency to avoid certain verb forms and structures. As evidenced by a study by Hinkel (2004), novice learners tend to avoid VVN forms due to their more complex sentence structure. It is also possible to hypothesize that the higher the proficiency of the writer, they are able to employ more variety of verb forms. This paper suggests that this hypothesis could be true to an extent, as shown by the findings of the variety of verb forms used. The population of this corpus is of advances ESL learners where they show no avoidance of any forms of verbs. They are also more partial to VVN forms, which involves complexity in structure but is generally considered academic prose convention. Although there appears to be underuse and overuse of certain verb forms, their use follows the genre rhetoric, where VVD forms are sparsely used. As such, the hypothesis is true to the extent that the frequency of verb forms are highly dependable to the rhetoric and conventions where in this case, avoidance of certain verb form does not necessarily indicative of level of proficiency but rather the intended meaning and function.

5.5 Tense and Aspect

Verb forms and lemmas are a source of great research analysis. Through verb forms, choice of tense, aspect and voice could be examined. It is acknowledged that present tense is more prevalent in academic writing. It is also true in this study where

University of Malaya

Present tense is employed more frequently than Past tense. The Past Tense is almost exclusively used in historical piece or biographical texts and they tend to appear in very small number in academic text. The occurrence of Past Tense in this corpus is considerably high in number thus further analysis is carried out. It is found that the writers favor narration as part of their discussion process. For example, significant occurrences (*the interviewee suggested that…)* during an interview is narrated in the discussion in order to create a cohesive discussion. This could have been avoided if the writer applies passive voice in place (*It is suggested that…).*

Choice of passive and active voice provides great research avenue. Many scholars have delved into this notion and a number previous research has arrived to contradictory conclusions. On the one hand, active voice is encouraged in academic writing due to its straightforward and precise nature. Contrarily, scholars consider passive voice to be the discourse convention. Passive voice enables writers to distance themselves from their writing, thus creating an illusion of objectiveness and detachment that shapes academic writing. It is suggested that novice writers, regardless of L1 background tend to create excessive writer visibly in their writing (Neff et al, 2004, as cited in Granger and Paquot, 2008). Writer's visibility is to be avoided as the subject matter should the main focus. Aside from creating an objective and indirect piece of academic prose, Passive voice also serves a textual function by forming cohesive links in sentences. This is achieved by using long passives. This method allows writers to use complex sentences while keeping the readers aware of the flow. Considering these arguments, writers should take deep consideration in which voice to use that is reflective of their intent. Previous literature has revealed that non-native writers tend to avoid using passive voice due to its more complex sentence structure. The fact in the present study finds the ESL writers in the corpus employ more passive voice renders the argument quite contentious. In any way, the findings could differ depending to many variable, one that include writers' proficiency,

University of Malaya

L1 background etc. It is more likely that novice writers tend to avoid passive voice while advanced writers, as are the population of this corpus, are able to apply both voices appropriately. For example, a study of Chinese learner by Hinkel (2002) noted that they face particular difficulty in constructing passive voice as the Chinese language does not have 'syntactically-derived' passive forms. Different voice may cater to specific discourse function, but in some cases they may very well interchangeable. It would be prudent for writers to equip themselves with knowledge and skills to apply both voice accordingly.

It is critical that the tense and aspect are employed consistently throughout the writing as consistency of tense and aspect generally the markers of highly proficient writers. However, example of inconsistent use of tense and aspect is found in this corpus which is generally a more prevalent problem among lower proficiency writers. In a study of Thai ESL writers, Pongsiriwet (2001) noted that inconsistency in verb structures, possibly caused by random change in verb structure for example inconsistent use of tense would cause understanding on the readers' part, as well as alter the intended meaning of the verbs. These elements would undoubtedly cause irreparable damage to the overall quality of the writing. With regards to this scenarios, it is suggested that ESL writing instructors teach and have students practice making natural shifts and flow of the verb structures so as to make their writing more comprehensible.

5.6 Common verb patterns

Language patterns or phraseology has been studied extensively in recent years, due to its pervasive and salient nature. Learning language in chucks is seen as a more applicable method in language acquisition. In academic writing notably, extensive research, especially corpus-based has learned that language is more often appears in

University of Malaya

phrases and collocates, thus led to compilation of academic collocations. In the scope of this study, Biber's classification of verb-based bundles are employed, although after much deliberation (frequency of usage, focus of study), only two are examined; *Passive + Preposition* fragment and *Verb + that* construction.

It is very difficult to apply general patterns in use of EAP verbs. Although there are notably growing number of studies focusing on the lexico-grammatical patterning of EAP vocabulary, they appear to be lacking suitability (Granger, 2006). On the one hand, some studies are very specific in their focus, such as study of colligation in specific field of ESP. On the other, there have been studies which identifies very broad number of linguistic patterns, such as patterns of tenses. As a result, learner and instructors alike are not able to apply the findings from these research pedagogically. In this manner, they lose their real-life application. Another point to note, as remarked by Granger (2006) is that phraseological studies have mostly focused on analysis of native-speakers use which another researcher has also voiced concern over. Flowerdew (1998, as cited in Granger, 2006) proposes that pedagogical implications of a study need to be in correspondence to the corpus population under analysis. This allows for adequate understanding of the population weakness or inadequacy where if measured properly, will only cause further problems. An example of possible scenario arising from mismatch implication is emphasis on teaching frequent patterns which learners are already familiar with, thus encouraging overuse on the learners' part.

It is also proposed by Granger (2006) that learners use of verbs in academic writing are not mainly distinguished by the choice of verbs per se, but rather the combination of patterns in which the verbs are placed in. Studies have found that in some learner's corpus, (Nesselhauf, 2005; Granger, 2006) the lexico-grammatical patternings of highly salient verbs appears to quite distinctive compared to their native counterpart particularly in

University of Malaya

active-passive alternation. Novice writers specifically are prone to use verbs in pattern. It is important to bear in mind that these verbs have their own 'preferred lexico-grammatical company'. For example, stock phrases for report verbs are almost consistently paired with that. Table 5.2 highlights these examples.

Table 5.2 Stock Phrases for Report Verbs

	Report verbs	
Smith' study	Proposes	that
	Suggest	
	Argues	
	Reports	

(Adapted from Hinkel, 2004)

As discussed earlier, the writers in the corpus are more partial towards passive voice. Passive voice is used with various prepositions with combination of *Passive + in* being the most frequently used (*as in cited in*).

(VP) + that clause fragment occurs in relatively high frequency in the corpus. It is typically used in as reporting verb, as suggested by previous literature. Selection of verb in (VP) + that clause fragment could be interpreted in many directions. One of them is the link between choice of verbs as epistemic device. It is proposed that choice of verb could indicate the level of confidence (*certainty* vs *likehood*) of the writers in relation to their statements. Readers can also employ this knowledge to gauge writers' feeling in regards to their choice of verbs. The study finds that certainty verbs occur more pervasively, although not by much. It is pointed out by Hyland (2002) the soft sciences are more partial to use likelihood verbs compared to the hard sciences. This is due to the patterns of reporting in the soft science in which findings need to be linked to previous studies. No two studies are the same thus arguments could only be made by persuasion.

University of Malaya

Such nature that is related to the discipline, and the compounding factor of non-native writers have produced the unlikely verbs that are *feel* and *believe*. The present study has tried to investigate the link between the choice of Epistemic verbs and the source of report but no clear link has emerged. At first it would largely appear that the certainty verbs such as *show* may be exclusively paired with inanimate subject (*data, finding* etc) but findings based on the verb *suggest* deny that particular conclusion. However, it is prudent to conclude that while there are examples of likelihood verbs employed with inanimate subject, the combination of certainty verbs and inanimate subjects is more prevalent than the other way around.

5.7 Summary of discussion

The most common verbs employed in the corpus is the Activity verb, followed by Mental-Emotive, Report, Logical Semantic and lastly Linking Verb. The higher occurrence of Emotive verbs raised concerns over the objectivity of academic prose. This finding, however, is also reflected in other studies where non-native writers are found to be more partial to use Emotive verbs. Analysis of verb forms allows us to see the tense and aspect favoured by the corpus population. In general, the corpus maintains the conventionalized use of passive voice in an effort to achieve academic detachment and objectivity. This finding is contrary to previous studies where they suggest that non-native writers tend to avoid complex sentences that characterize passive sentences. Regardless, overuse of passive voice could signal certain concerns when use of active voice more promoted for academic constructs that makes for more concise and profound read. Lastly, are identified as the most common verb-based patterns in the corpus. As discussed earlier, passive voice is the familiar voice in academic writing. Choice of preposition paired with passive voiced are relatively varied with highest combination with preposition *in* and *by*.

University of Malaya

These two prepositions are most commonly used to refer to previous studies or the author themselves

.

5.8 Pedagogical Implications

The findings from the study could pinpoint to certain deficiencies in the current situation that needs further improvement. Note that these implications are based on analysis of small and very specific population of a corpus thus may not reflect the general scenario.

5.8.1 Academic Wordlist

Compilation of high-frequency academic words is has proven to be quiet in trend in recent years to cater to developing and expanding English academic writers. One of the most prominent academic wordlist is one created by Averil Coxhead for her M.A thesis in 2000. Others include Pearson Longman Academic Word List and New Academic Word List. The Coxhead Academic Word List consists of 570 word families which are chosen based on frequency in the Academic corpus, across all academic fields. The list is designed to exclude most common English words. This method of compilation allows author to exclude words that are highly salient in daily conversation but hardly present in academic writing. The primary purpose of the Academic Wordlist is to serve students at tertiary level where the words are highly associated with academic learning and writing. Similar wordlists are also very popular among learners where learner find them very useful to assist them in their academic pursuit. The list however does not provide indication of word category membership, for example words such as *conduct, present* or

University of Malaya

study which can function as verbs and nouns. Therefore, it is critical that these lists include word categories or perhaps a list of academic verbs are specially compiled.

In response to the aforementioned issues, this paper is an attempt to address them in a pragmatic manner. One underlying characteristic of this this study is that it has chosen a specific population in mind. In this case, the writers are highly immersed in the academic setting, having to write up academic papers and dissertations. They also make for valuable research focus in which the results are likely to address some of the issues they are having. Also, by focusing on more generic verb types and forms that are derived from the corpus itself, the findings will be more useful and reflective of the corpus. In this manner, this study distinctive from other studies whereby the focus is often solely on highly frequent EAP words. Having discussed this matter at length earlier, the most frequent verbs found in the corpus do not necessarily coincide with the academic verb lists (*believe, feel*). These sort of finding will prove to be invaluable source of examination in an effort to understand and eliminate the problem. The process of compiling academic verbs involves tenuous process of analyzing and categorizing large data. With the advancement of technology, larger data can be processed thus ensuring reliable results that are applicable to specific genre. However, there are certain issues the need fixing in order for the results to make great impact. As discussed previously, it is not a matter of lack of effort in compiling the most comprehensive EAP guideline, but there is a small problem in adjusting the balance of focus of the research so that they could make greater impact pedagogically. Insufficient knowledge of EAP verbs is a serious handicap for learners as it prevents them from expressing their thoughts in all their nuances and couching them in the expected style. By exposing the list of lexical verbs or EAP verbs to the learners and writers, their thoughts and emotions and predicate meanings can be conveyed into written text and consequently without a doubt, an imperative first step, but unless it is set off with a detailed depiction of their use, outcomes are bound to be extremely unsatisfactory.

University of Malaya

Furthermore, the use of the list could benefit learners in avoiding repetition, as exhibited from the findings of this paper as well as previous studies.

5.8.2 Teaching of Grammar

The present study has noted some grammar inaccuracy and inconsistency that occurs even in advanced ESL writing. This occurrence has raised the of the old age question of whether grammar teaching is still applicable for ESL learners at tertiary education level or more specifically advanced ESL. It has been remarked by previous researchers that many ESL instructors have more concerns over content and structure of academic writing instead of going back to the basics of grammar which is assumed to have been exposed earlier in their L2 education. However, it is also safe to assume that some advance ESL learners might have received less than adequate of exposure in grammar rules, especially for academic genre which is notably more intricate and complex.

Bearing in mind of these dilemmas, it is prudent to conclude that grammar should be taught in accordance to the genre, in that specific curricula is tailored to the needs and requirements of academic genre. This paper suggests that grammar is viewed as part of the content and organization of an actual academic writing task.

As discussed in RQ 2, usage of tense and aspect of verbs in academic writing is reflection of the writers' ability to modulate multiple meanings and functions, and choose the most discourse accurate verb forms. The outcomes of these complicated process is a coherent and cohesive piece of writing. In other words, teaching English verb tense and aspect in this case would not refer to simply teaching the grammatical concepts of each verb structure by defining each form, listing its usages, or having practice drills for

University of Malaya

multiple choice questions. Rather, it would be linking verb usages to the discourse-level features and the content of the essay so as to help students build up logical and well-organized ideas in a coherent manner.

Familiarity with discourse-specific grammatical features will enable learners to create more sound and logical arguments, while avoiding ambiguous and vague structures. The open secret to achieve consistency in writing lies in the way the relationship is built between idea which involves modulation of verbs specifically. Witte & Faigley (1981) claims writing quality to depend greatly on the discourse-level features that "lie beyond sentence boundaries involving the underlying relations between ideas which allow a text to be understood" (as cited in Pongsiriwet, 2001, p.87). Therefore, it would be greatly beneficial for learners to be explicitly taught discourse-level verbs in terms of time, aspect and tone as they could shape the cohesive and coherentness of the end product. Thus, giving explicit instructions on understanding and using English verb tense and aspect would not be "teaching grammar" but would rather be a practical application of integrating grammar with content and structure which will improve NNS students' writing proficiency.

5.8.3 Explicit teaching of Phraseology and Collocation

Previous chapter has described the enormity and ubiquitous nature of phraseology, and that there is no disputing the fact that it is essential in the process of language learning. As Gitsaki (1996) stated in her thesis, the importance of phraseological aspect of linguistics development and communicative competence for L2 learners have long been underscored by linguists and language teachers alike, very much so that they are endorsing teaching and learning of salient language patterns and collocations in a language classroom. On the other end of the spectrum, it is suggested

University of Malaya

that non-native learners face particular difficulty in producing appropriate word combinations because of their lack of collocational knowledge (Howarth, 1998). An empirical study on the knowledge of collocations among different groups of ESL or EFL learners have confirmed their dilemma. A study by Li (2005) in a Taiwanese college of EFL context found inconsistency in the collocational errors made and their perceived knowledge of collocations. Liu (1999b) found that the EFL students had difficulties in producing correct collocations is attributed to the lack of the concept of collocation. Insufficient knowledge of collocation encourages learners to resort to strategies such as synonym, paraphrasing, avoidance, and transfer (Farghal and Obiedat, 1995). Fan (2009) claims that the 'idiosyncratic' nature of collocation is the cause for errors in ESL learners. She cited Halliday's (1996) example that 'strong' and 'powerful' are synomymous, one can only refer to 'tea' and 'strong' and not 'powerful tea'.

Consequently, on a separate level, collocation has been considered as a separate level of vocabulary acquisition. Bolinger (1976) argues that we learn and memorise words in chunks and that most of our "manipulative grasp of words is by way of collocations". It has been argued that the teaching of collocations facilitates vocabulary building for University-bound ESL students (Smith 1983). Smith (1983), in his paper illustrates a type of exercise for the teaching of collocations that combines both paradigmatic and syntagmatic relations between words. Realising the foreign language learner's difficulties in learning vocabulary, Cowie (1978) stresses the importance of the compilation of English dictionaries "in which collocation and examples play a separate but complementary role". Cowie points out that "meaning is not the only determinant of the extent and semantic variety of collocating words.... The constraint may be situational". Here he emphasized the role of semantic elements are not the only determining factors that define collocation but rather *a factor* among many others. A such, inclusion of

University of Malaya

collocational items in the dictionary may provide for a wholesome learning experience for ESL learners.

5.8.4 Consciousness-raising activity via data-driven learning (DDL)

A contrary approach to teaching of phraseology could also be a handy approach for teaching in learning. It has been discussed preciously that explicit teaching and learning of phraseology could be a way to go but one could consider employing an implicit and subtle method in introducing phraseology through DDL.

One of the benefits of using corpus-based materials is discussed by Hunston and Francis (2000, p. 268), "If the learner wishes to sound 'natural', 'idiomatic', or 'native-like', it is argued, he or she needs to use the collocations, the phraseologies and the patterns of English that native speakers automatically choose". The knowledge of collocations has been widely recognised as an important aspect in language learning (Howarth, 1998; Nation, 2001). They agreed that the "appropriate use of collocations enables the learners to speak more fluently, makes their speech more comprehensible and helps them produce more native-like utterances" (as cited in Hong et al., 2011, p. 2) and therefore plays a very important role in Second Language Acquisition (Sinclair, 1991; Howarth, 1998; Nation, 2001; Durrant & Schmitt, 2010). Cortes (2004), for example, notes that the proficient language use is marked by use of collocations and fixed expressions, and approvingly quotes Haswell's (1991) claim that as writers grow more mature in terms of their writing, their use of collocations will grow in numbers as well. Similarly, in their studies of the development of collocational knowledge in non-native writers, both Nesselhauf (2005) and Kazsubski (2000) assume that increased proficiency will correlate with increased use of conventional collocations. It is established that native speakers share a substantial body of formulaic sequences thus it is the second language speaker's ability to gain

University of Malaya

collocational knowledge that determines, in part, language learning success. Granted, achieving sufficient collocational knowledge is not as simple as we would like to imagine. In addition, there are certain verb structures that are more complex than the others, which are also difficult to teach in a traditional classroom. Language environment is also one of the elements that separates non-native speakers' ability in utilising collocations to their native counterpart. Native speakers are able to apply the collocational knowledge easily as they recognized the formulas as unit with clearly defined functions, due to vast exposure of the target language (Ellis et al., 2008). Contrarily, non-native speakers, less exposed to the language, flounders in using collocations. Non-native can only absorb this knowledge if they were immersed in a speaking community where they can observe and imitate (Wray, 2002).

At this juncture, it appears that DDL approach to phraseology could be the revolutionary step forward. Bearing the learners' as well as instructors' needs in mind, researchers have been looking this aspect of language that could help non-natives achieve similar competency as their native counterparts. Previous studies have concurred the suggestion that the frequency-driven formulaic expressions found in native expert writing can be of great help to learner writers to achieve a more native-like style of academic writing, and should thus be integrated into ESL/EFL curricula. A study by Ellis et.al (2008) further reaffirms the benefits of formulaic sequences whereby he also advocates the use of academic language formulas in EAP instructions. In his study, he triangulates data from three different perspectives; psycholinguistics, education and ESL. Through that study, he successfully proves that the study of formulaic language is relevant from all three perspectives. In response to these findings, consciousness-raising activity can be beneficial especially for advance learners in improving their collocational knowledge. Consciousness-raising' is an important element in creating awareness with regard to collocations and this has been widely acknowledge and suggested by many previous

University of Malaya

researchers (Howarth, 1996; Hill et al., 2000; Lewis, 2000; Woodlard, 2000). This type of activity is especially valuable in assisting learner to actively note and evaluate the forms and functions of lexical bundles and allow them to use the phrases in their own writing accurately and effectively. For novice writers, these activities can help them to notice the similarities and differences of these bundles to their first language, and avoid making errors related to interference of their L1. Use of corpora and concordance to assimilate learners to 'real-worlds' language use has been experimented on for many years and results have indicated positive developments towards teaching and learning of writing (Lu, 2002). In his study, it is noted that that subjects responded that DDL had improved their knowledge of the form, meanings, and actual usages of legal English sentences. DDL increased their knowledge of prepositional rules and the inductive learning activities helped them become more independent students.

Finding from previous studied have suggested that DDL method have helped them in increasing their awareness in the variety of meanings and usages collocations and colligations. The specific way that this is achieved is through explicit learning of concordance lines where their noticing skill is tested. These findings have led researchers such as Ellis et al. (2008) and Wray (2002) to remark the sustainability for research in these areas. However, currently there is still lack of notable research emerging from this area, particularly ones that provide empirical evidence to encourage further study. Another researcher, Koosha and Jafarpour (2006) pinpoints the need of the study of phraseology and collocation to have tight link with L1. In other words, comparative study with L1 should be reference point for any studies in this area. Present study has taken this into consideration at earlier stage of research but it was not encouraged. However, as showed elsewhere in this study, this researcher does not deny the fact that some other factors, for example task types, learner proficiency, and level of education still play an influential role in influencing complete acquisition of phraseological knowledge.

University of Malaya

5.9 Conclusion

This study has focused on lexical verb use among ESL advanced learners in a university in Malaysia. The results from this study highlights a few implications for EAP classrooms as well as suggestions for future research. The first suggestion that this paper would like to put forth to encourage the use of academic wordlists among L2 learners, particularly when they are writing their papers. There are a number of established academic wordlists, which are the results of rigorous research and analysis. The wordlists may be a general list that cover all fields of study or it could be a specific one that are tailored for certain field or chapters in an academic paper. The lists have been proven as a handy tool to have and use in EAP classrooms or during the process of writing. Another implication is a suggestion for EAP instructors and language teachers to consider going back to teaching grammar in language classrooms. They must not forget that some of their learners maybe lacking exposure to the proper teaching of grammar, especially the genre-related aspects. The research for the past twenty years have pointed out the importance of phraseology in writing. This research also highlights the importance of collocations in academic writing and proposes analysis of common patterns as linguistic awareness exercise for advanced ESL learners. The teaching and learning of phraseology can be approached in two contradictory manners, explicit and implicit teaching. Both methods have gained positive feedbacks from EAP communities especially the DDL method which could shape the pedagogical approach in EAP classrooms particularly and teaching and learning of language in general.

Although the Malaysian academic community is slowly gaining momentum in conducting research in corpus linguistics, it also seems that we are falling behind compared to other nationalities such as the Turkish, Chinese and German scholars.

University of Malaya

Considering the huge benefits from corpus studies as well as the gap in academic input from the Malaysian scholars, this situation needs to be remedied soon. More importantly, taking into account the evident importance of verbs in academic writing, it is imperative that further study is conducted to further illustrate the characteristics of lexical verbs in academic discourse in Malaysia.

University of Malaya

REFERENCES

Agçam, R. (2015). Author stance in academic writing: A corpus-based study on epistemic verbs. *The Journal of Teaching English for Specific and Academic Purposes* Vol. 3, No 1, 2015, pp. 9-20.

Altenberg, B. (1998). *On the phraseology of spoken English: the evidence of recurrent word-combinations.* Oxford: Oxford University Press.

Altenberg, B. & Granger, S. (2001). The Grammatical and Lexical Patterning of MAKE in Native and Non-Native Student Writing. *Applied Linguistics*, 22 (2), 173.

Ang Leng Hong, Hajar Abdul Rahim, Tan Kim Hua, Khazriyati Salehuddin. (2011). Collocations in Malaysian English learners' writing: A corpus-based error analysis, 3L. *The Southeast Asian journal of English language studies*. Vol. 17(1), 31-44.

Duffley, P. & Arseneau, M. (2012). Tense and control interpretations in gerund-participle and to-infinitive complement constructions with verbs of risk. *Canadian Journal of Linguistics/Revue canadienne de linguistique.* 57. 31-50.

Barton, D. (1994). *Literacy: An introduction to the ecology of written language*. Oxford, UK: Blackwell.

Bahns, J. (1993). Lexical collocations: a contrastive view. *English Language Teaching Journal*, 47(1), 56-63.

Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocation? *System*, 21(1), 104-114.

Balikçi, Gözde (2011) *The Use Of Collocations By Advanced Learners Of English: Noun –Noun And Adjective-Noun*
Collocations. In: 1st International Conference on Foreign Language Teaching and Applied Linguistics (FLTAL'11), 5-7 May 2011, Sarajevo.

Benjamins, J. (2014). Corpus-informed research and learning in ESP. Issues and applications. *Iberica*, 27: 217-234.

Bennet, A. (1991). *Discourses of power, the dialectics of understanding the power of literacy.* New York: Bergin and Garvey.

Bernard, A. M. and Winnie, A. Y. L. (1985). Academic writing and Chinese students: Transfer and developmental factors. *TESOL Quarterly*, 19(3): 515-534.

Benson, M., Benson, E. & Ilson, R. (1986). *Lexicographic description of English.* John Benjamins Publication Company.

Bhatia, V.J., Sánchez Hernández, P. & Pérez- Paredes, P. (2011). *Researching Specialized Languages*. Amsterdam: John Benjamins.

University of Malaya

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D. and Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for specific purposes,* 26: 263-286.

Biber, D. and Conrad, S. (2001). *Variation in English: Multi-dimensional studies*. London: Pearson.

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

Biber, D., Conrad, S. and Reppen, R. (1998). Corpus linguistics: investigating language structure and use. Cambridge: University Press.

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

Biber, D. & Conrad, S. (2009). *Register, Genre and Style*. Cambridge: Cambridge University Press.

Biskup, D. (1992). L1 influence on learners' renderings of English collocations: a Polish/German empirical study'in PJL Arnaud and H. BeÂjoint (eds.): Vocabulary and Applied Linguistics. *Vocabulary and Applied Linguistics*. 85-93.

Bloch, J. (2010). A concordance-based study of the use of reporting verbs as rhetorical devices in academic papers. *Journal of Writing Research,* 2 (2), 219-244

Bolinger, D. (1975) *Aspects of Language*, New York: Harcourt Brace Jovanovich.

Boulton, A. & Tyne, H. (2013). Corpus linguistics and data-driven learning: A critical review. *Bulletin Suisse de linguistique appliqué*, 97: 97-118.

Bowker, L. and J. Pearson (2002), *Working With Specialized Language: A Practical Guide to Using Corpora*. London: Routledge.

Burrough-Boenisch, J. (2003). Editing texts by non-native speakers of English. *Science Editors' Handbook.* 1-4.

Cadman, K. (1997). Thesis writing for international students: A question of identity?. *English for Specific Purposes,* 16(1): 13-14.

Callies, M. (2013). High-frequency nouns and their verbal collocates in academic learner writing. *Learner Corpus Research 2013*, 27-29 September 2013, Bergen, Norway.

Callies, M. & E. Zaytseva (2013) The Corpus of Academic Learner English (CALE) – A new resource for the assessment of writing proficiency in the academic register. *Dutch Journal of Applied Linguistics* 2(1): 126-132.

Charles, M. (2006). Phraseological patterns in reporting clauses used in citation: a corpus-based study of theses in two disciplines. *English for Specific Purposes*, 25: 310-331.

University of Malaya

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

Chomsky, N. (1965) *Aspects of the Theory of Syntax*. The MIP Press.

Citravelu, N. Sithamparam, S. and Teh, S. C., 2005. *ELT methodology: principle and practice*. 2nd ed. Shah Alam: Penerbit Fajar Bakti Sdn. Bhd.

Clarke, P. (2009). *Supporting children learning English as a second language in the early years.* Victorian Curriculum and Assessment Authority.

Conti, G. (2011). Defining a Rule for the Use of Infinitive and Gerund Complements. *English Language Teaching* Vol. 4:3.

Cortes, V. (2002). *Lexical bundles in freshman composition*. Amsterdam: John Benjamins Publishing Company.

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

Cox, Christopher (2011). Corpus linguistics and language documentation: challenges for collaboration. In John Newman, R. Harald Baayen, & Sally Rice (eds.), C*orpus-based studies in language use, language learning, and language documentation,* 239-264. Amsterdam: Rodopi.

Coxhead, A. (2000). A New Academic World List. *TESOL Quarterly*, 34(2): 213-238.

Crystal, D. (1997). *A dictionary of linguistics and phonetics.* Oxford: Blackwell. Fourth Edition.

Crystal, D. (2003). *The Cambridge Encyclopaedia Of The English Language*. Cambridge: Cambridge University Press.

Darmi, R., & Albion, P. (2013). *English Lanuguage in the Malaysian Education System: Its Existence and Implications*. 3rd Malaysian Postgraduate Conference, Sydney, New South Wales, Australia, (July), 175–183.

Davies, B. and Harre, R. (1990). Positioning: The discursive production of selves. *Journal for the Theory of Social Behavior*, 20(1): 43-63.

De Cock, S., Granger, S., Leech, G. and McEnery, T. (1998). *An automated approach to the phrasicon of EFL learners.* London: Longman.

Durrant, P. (2008). *High frequency collocations and second language learning.* (Unpublished doctoral dissertation). The University of Nottingham, Nottingham.

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, 26(2), 163-188.

University of Malaya

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second-language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 41(3), 375-396.

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

Fan Shen. (1988). The classroom and the wider culture: identity as a key to learning English composition. *College Composition and Communication,* 40: 459-466.

Farghal, M., & Obiedat, H. (1995). Collocations: A neglected variable in EFL. *IRAL*, 33(4),315-333. Retrieved 15 November, 2013 from http://dx.doi.org/10.1515/iral.1995.33.4.315

Firth, J. R. (1957). Modes of meaning. In *Papers in linguistics* 1934-1951 (pp. 190 -215). Oxford: Oxford University Press.

Foster, P. (2001) "Rules and Routines: A Consideration of their Role in the Task -Based Language Production of Native and Non-Native Speakers" in Bygate, M., Skehan, P. and Swain, M. (Eds.) *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing,* London: Longman

Folse, K. S. (2004). *Vocabulary myths: Applying second language research to classroom teaching*. Ann Arbor, MI: University of Michigan Press.

Francis, N. and Kucera, H. (1982). *Frequency analysis of English usage: lexicon and grammar*. Boston: Houghton Mifflin.

Gardner, D. and Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus based analysis. *TESOL Quarterly*, 41(2): 339-359.

Gass, S. M. and Selinker, L. (2001). *Second language acquisition: an introductory course* (2nd Eds.). Hillsdale, NJ: Lawrence Erlbaum.

Gee, J. P. (2005). An introduction to discourse analysis: *Theory and method*. (2nd Ed). New York: Routledge.

Gibson, M. (2005). Promoting academic engagement among minority youth: Implications from John Ogbu's Shaker Heights ethnography. *International Journal of Qualitative Studies in Education*, 18(5): 581-603.

Gitsaki, C. (1996) *The development of ESL collocational knowledge.* Unpublished PhD Thesis. University of Queensland.

Gobel, P., Thang, S. M., Sidhu, G. K., Oon, S. I., & Chan, Y. F. (2013). Attributions to success and failure in English language learning: A comparative study of urban and rural undergraduates in Malaysia. *Asian Social Science,* 9(2), 53–62.

Goldberg, A. E. (2013). Argument Structure Constructions versus Lexical Rules or Derivational Verb Templates. *Mind & Language*, 28, 435–465.

Granger, S. (Ed.) (1998) *Learner English on Computer*. Longman.

University of Malaya

Granger, S. and Meunier, F. (2008). *Phraseology: An interdisciplinary perspective*. Amsterdam: John Benjamins.

Granger, S. and Paquot, M. (2005). *The phraseology of EFL academic writing: Methodological issues and research findings.* Paper presented at AAACL6 & ICAME26, 12-15 May 2005, Ann Arbor, Michigan.

Granger, S. and Paquot, M. (2008). Disentangling the phraseological web. In: Granger, S. and Meunier, F. (Eds.)., *Phraseology: An interdisciplinary perspective.* Amsterdam: John Benjamins.

Granger, S. and Paquot, M. (2015). Lexical verbs in academic discourse: A corpus-driven study of learner use. In: H. Basturkmen. *English for Academic Purposes.* London: Routledge.

Gries, Stefan. (2008). *Phraseology and linguistic theory: A brief survey*. 3-25

Gries, S. Th. & Stefanowitsch. A. (2004a). Extending collostructional analysis: A corpus-based perspectives on 'alternations'. *International Journal of Corpus Linguistics* 9.1:97-129.

Halliday, M.A.K. (1989) *Language, Context, and Text: aspects of language in a Social-semiotic Perspective.* Oxford University Press

Halliday, M.A.K. (1994). *Introduction to Functional Grammar*, 2nd edition, London: Edward Arnold.

Hardie, A. & McEnery, T. (2010), On two traditions in corpus linguistics, and what they have in common. *International Journal of Corpus Linguistics*, 15: 384-394.

Haswell, R. (1991). *Gaining ground in college writing: tales of development and interpretation*. Dallas: Southern Methodist University Press.

Henriksen, B. (2012). Research on L2 learners' collocational competence and development – a progress report. In Bardel, C., C, Lindqvist, & B. Laufer (Eds.). *L2 vocabulary acquisition, knowledge and use. New perspectives on assessment and corpus analysis* (pp. 29-56).

Hinkel, E. (2002). *Second language writers' text. Linguistic and rhetorical features.* Mahwah, New Jersey and London: Lawrence Erlbaum Associates.

Hinkel, E. (2004). *Teaching academic ESL writing: Practical techniques in vocabulary and grammar*. Mahwah, New Jersey and London: Lawrence Erlbaum Associates.

Howarth P. (1996): *Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making*, Tübingen

Hunston, S. (2000). *Evaluation and the planes of discourse: Status and value in persuasive texts*. Oxford: Oxford University Press.

University of Malaya

Hunston, S., & Francis, G. (2000). *Pattern grammar. A corpus-driven approach to the lexical grammar of English.* Amsterdam: John Benjamins

Huddleston, Rodney and Geoffrey K. Pullum (2002). *The Cambridge Grammar of the English Language.* Cambridge: Cambridge University Press.

Hyland, K. (1998). Boosting, hedging and the negotiation of academic knowledge. *TEXT* 18 (3) pp. 349-382.

Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics*, 20: 341-367.

Hyland, K. (2000). *Disciplinary Discourses: Social interactions in academic writing*. London: Longman.

Hyland, K. (2002a). Options of identity in academic writing. *ELT Journal*, 56(4): 351-358.

Hyland, K. (2002b). authority and invisibility: Authorial identity in cademic writing. *Journal of Pragmatics*, 34: 1091-1112.

Hyland, K. (2005). Representing readers in writing: Students and expert practices. *Linguistics*, 16(4): 363-377.

Hyland, K. (2007). Applying a gloss: Exemplifying and reformulating in academic discourse. *Applied Linguistics*, 28(2): 266-285.

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

Hyland, K. (2013). Writing in the university: Education, knowledge and reputation. *Language Teaching*, 46(1): 53-70.

Hyland, K. and Milton, J. (1997). Qualifications and certainty in L1 and L2 students writing. *Journal of Second Language Writing*, 6(2): 183-205.

Hyland, K. and Tse, P. (2007). Is there an Academic Vocabulary? *TESOL Quarterly*, 41(2): 235-253.

Ishikawa, S. (2013).The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. *Learner Corpus Studies in Asia and The World* 1:91-118.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, 21(1): 47-77.

Jolita Horbačauskienė & Saulė Petronien. (2013). Verb Valency Patterns in Academic Register: Syntactic Approach. *Studies about Languages* 22, 11-16.

University of Malaya

Kamariah Yunus & Su'ad Awab. (2011). Collocational competence among Malaysian Law undergraduate students. *Malaysian Journal of ELT Research*,7(1),151-202. Retrieved 15 November, 2013 from http://www.melta.org.my/Volume-7-1-2011/q?cid=93&doit=showclass

Kaplan, R. (1972). *The anatomy of rhetoric: prolegomena to a functional theory of rhetoric.* Philadelphia: Center for Curriculum Development. (Concord, Massachusetts: Heinle and Heinle).

Katnic-Bakarsic, M. (2004). Discouse stylistic as contextualized stylistics. *Journal for Philogical Research*, 15(2): 37-48.

Kennedy, G. (1998). *An introduction to corpus linguistics.* London: Longman.

Knowles, G. and Zuraidah, M. D. (2004). The notion of a lemma: Headwords, roots and lexical sets. *International Journal of Corpus Linguistics,* 9(1): 69-81.

Lea, M. & Street, B. (1999). *Writing as academic literacies: understanding textual practices in higher education.* Harlow: Longman.

Leech, G. (1992). *Corpora and theories of linguistic performance.* In J.Svartvik (ed.). 105–122.

Leech, G. N., Rayson, P. and Wilson, A. (2001). *Word frequencies in written and spoken English.* London: Longman.

M. Koosha & A. Jafarpour.(2006). Data-driven learning and teaching collocation of prepositions: The case of Iranian EFL adult learners. *Asian EFL Journal,* 8 (8), pp. 64-82

Maroko, G. M. (2013). Learning about author positioning in written academic discourse. *Argentinian Journal of Applied Linguistics,* 1(2): 47-60.

Matsuda, P. K. & Tardy, C. M. (2007). Voice construction in academic writing: The rhetorical construction of author identity in blind manuscript review. *English for Specific Purposes*, 26(2): 235-249.

McEnery,T. & Costas, G. (2006). English corpus linguistics. In B. Aarts and A. McMahon (eds.). The handbook of English linguistics, 33–71.Oxford: Blackwell.

McEnery, T., & Wilson. (2001). *Corpus Linguistics.* Edinburgh: Edinbugh University Press.

McEnery, T., Xiao, R. & Tono, Y. (2006). *Corpus-based Language Studies: An Advanced Resource Book.* London: Routledge.

Meyer, P. G. (1997). *Coming to know: studies in the lexical semantics and pragmatics of academic English.* Tubingen: Gunter Narr Verlag Tubingen.

University of Malaya

Min, Kyung Eun. (2013). *How grammar matters in NNS academic writing: the relationship between verb tense and aspect usage patterns and L2 writing proficiency in academic discourse.* (Unpublished Master's Thesis). University of Illinois at Urbana-Champaign.

Murugesan, V. (2003). Malaysia Promotes Excellence in English. *ESL Magazine*, (March/April), 26–28.

Myers, G. (1990). *Writing Biology: Texts in social construction of scientific knowledge.* Madison: University of Wisconsin Press.

NA Manan, NM Noor (2014). Analysis of Reporting Verbs in Master's Theses. *Procedia-Social and Behavioral Sciences* 134, 140-145

Neff, J., Ballesteros, F., Dafouz, E., Martinez, F. and Rica, J. P. (2004). The expression of writer stance in native and non-native argumentative texts. In: R. Facchinetti. And F. Palmer (Eds.), *English modality in perspective*. Frankfurt am Main: Peter Lang. (pp.141-161).

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223-242.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: Benjamins.

Oakey, D. J. (2005). Academic vocabulary in academic discourse: The phraseological behavior of EVALUATION in Economics research articles. In: E. Tognini-Bonelli and G. Del Lungo Camiciotti (eds.), *Strategies in Academic Discourse.* Amsterdam: Benjamins.

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

Paquot, M. (2007). *Exemplification in learner writing: A cross-linguistic perspective.* Amsterdam: Benjamins.

Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics,* 24: 146-161.

Rohter, L. (2004). Learn English, says Chile, thinking upwardly global. *New York Times,* (pp. A3).

Russell, K. M. (2015). *A Comparison of Linguistic Features in the Academic Writing of Advanced English Language Learner and English First Language University Students.* (Unpublished Master's Thesis) Portland State University

Sadeghi, Karim and Shirzad Khajepasha, Arash (2015). Thesis writing challenges for non-native MA students. *Research in Post-Compulsory Education,* 20, 357-373. Routledge

University of Malaya

Safnil Arsyad. (2013). A Genre-Based Analysis on Discussion Section of Research Articles in Indonesian Written by Indonesian Speakers. *International Journal of Linguistics* Vol. 5, No. 4

Schmitt, N. (2004). *Formulaic sequences: acquisition, processing and use*. Amsterdam: John Benjamins.

Schutz, N. 2013. How Specific is English for Academic Purposes? A look at verbs in business, linguistics and medical research articles. In G. Andersen and K. Bech (eds.). *English Corpus Linguistics: Variation in Time, Space and Genre*. Amsterdam: Rodopi Publishers: 237-257.

Stubbs, M. (2001). *Words & Phrases: Corpus Studies of Lexical Semantics.* Oxford, UK: Blackwell.

Sinclair, J. (1987). *Looking up: an account of the COBUILD project in lexical computing*. London and Glasgow: Collins.

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford: Oxford University Press.

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review*, 64(3), 429-458.

Souriyavongsa, T., Rany, S., Abidin, M. J. Z., & Leong, L. M. (2013). Factors Causes Students Low English Language Learning: A Case Study in the National University of Laos. *International Journal of English Language Education*, 1(1), 179–192.

Stapleton, P. (2002). Critiquing voice as a viable pedagogical tool in L2 writing: Returning the spotlight to ideas. *Journal of Second Language Writing*, 11(3): 177-190.

Stubbs, M. (1996). *Text and corpus linguistics.* Oxford: Blackwell.

Stubbs, M. (2001). *Words & Phrases: Corpus Studies of Lexical Semantics.* Oxford, UK: Blackwell.

Wang, Y., & Shaw, P. (2008). Transfer and universality: Collocation use in advanced Chinese and Swedish learner English. *ICAME Journal*, 32, 201-232.

Stubbs, M. (2007). *Quantitative data on multi-word sequences in English: The case of word 'world'*. London: Continuum.

Swales, J. (1990). *Genre analysis. English in academic and research settings.* Cambridge: Cambridge University Press.

Swales, J. (2004). Then and now: A reconsideration of the first corpus of scientific English. *Iberica*, 8: 5-21.

Swales, J. and Feak, C. (2004). *Academic writing for graduate students. Essential task and skills*. Ann Arbour: The University of Michigan Press.

University of Malaya

Tabors, P. O. (1997). *One child, Two Languages: A guide for Preschool Educators of Children Learning English as a second language.* USA, Baltimore: Paul Brookes Publishing.

Tan, L. H., Hoosain, R. and Peng, D. (1995). Role of early presemantic phonological code inchinese character identification. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 21: 43-54.

Tang, R. and John, S. (1999). The "I" in identity: Exploring writer identity in student academic writing through the first person pronoun. *English for Specific Purposes,* 18(1): 23-39.

Teubert, W. (2005). My version of corpus linguistics, *International Journal of Corpus Linguistics,* 10: 1-13.

Thomas, S. and Hawes, T. P. (1994). Reporting verbs in medical journals. E*nglish for specific Purposes,* 13: 129-148.

Thompson, G. and Yiyun, Y. (1991), 'Evaluation in the reporting verbs used in academic papers'. *Applied Linguistics*, 12, (4), 365-382.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: Benjamins

Wallwork, A. (2011). English for writing research papers. *Springer Science+ Business Media.*

White, J. W. and Lowenthal, P. R. (2011). Academic discourse and the formation of an academic identity: Minority college students and the hidden curriculum. *Review of Higher Education*, 34 (2): 1-47.

Williams, I. A. (1996). A contextual study of lexical verbs in two types of medical research reports: clinical and experimental. *English for Specific Purposes,* 15(3): 175-197.

Wilson, A. (2000). *Conceptual glossary and index to the vulgate translation of the gospel according to John.* Hildesheim: Olms-Weidmann.

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

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

Wray, A. (2009). Future directions in formulaic language research. *Journal of foreign languages,* 32, 2-17.

Yannuar, Nurenzia et al. 2014. Active and Passive Voice Constructions by Indonesian Student Writers". *Theory and Practice in Language Studies,* Vol. 4, No. 7, pp. 1400-1408.

University of Malaya

Zhang, X. (2008). *A contrastive study of reporting in Master's Theses in native Chinese and in Native English. Dissertation presented for the degree of MSc in Applied Linguistics.* The University of Edinburgh.

University of Malaya