Malaysian learners’ argumentative writing in English: A contrastive, corpus-driven study

Siti Aeisha Joharry

A thesis submitted in fulfilment
of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics
University of Sydney

December 2016

Declaration

I certify that this thesis, submitted in fulfilment of the requirement for the award of Doctor of Philosophy in the Department of Linguistics, University of Sydney, does not incorporate without any acknowledgement or any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person where due reference is not made in the text.

Siti Aeisha JOHARRY

December 2016

Abstract

Research on learner English is by now an established sub-discipline in corpus linguistics, yet few studies exist on Malaysian learners. This thesis explores the difficulties that Malaysian learners of English face when producing argumentative essays, focussing on their overuse of particular linguistic features. WordSmith Tools (Scott, 2012) is used to analyse and compare two corpora: The Malaysian Corpus of Students' Argumentative Writing (MCSAW): Version 2, consisting of 1,460 Malaysian students' argumentative essays; and the Louvain Corpus of Native English Essays (LOCNESS), which is a corpus of native English essays written by British and American students and is used as a reference language variety here. The software enables analysis of keywords (words that are over-used in MCSCAW), collocates or surrounding words of the keywords, and concordances, which are used to examine the keywords in context. Crucially, it also allows examination of the 'range' of linguistic features (i.e. by how many students a feature is employed) – an under-used but crucial affordance of this software programme that is exploited in this thesis for down-sampling purposes. The thesis combines quantitative and qualitative corpus linguistic techniques, with keywords providing the starting point for in-depth qualitative analysis using concordancing.

This corpus-driven analysis of MCSAW identifies typical features of the writing style of Malaysian learners' writing of English, particularly the overuse of *can* and *we* (including the highly frequent bundle *we can*), and the lack of discourse-organising markers. Analysis of key words and key bundles is complemented with collocation analysis and concordancing of the highly frequent modal verb *can* as well as the highly frequent first person plural pronoun *we*, which both have a high range across the corpus. The concordances are carefully and systematically examined to explore the ways in which these over-used linguistic items are actually employed in their co-text by the Malaysian writers. While results show some similarities in both learner corpus and reference language variety, Malaysian learners tend to demonstrate higher writer visibility overall. One possible explanation lies in the influence of the national language (Malay). The thesis also identifies repeated sentences that occur in more than one essay, which implies either plagiarism on the learners' part or a particular teaching strategy (templates or phrases that are provided to students). This finding has significant implications for corpus design (in terms of the need for more topic variation) as well as methodological significance (in terms of the advantages and disadvantages of using the 'range' feature for down-sampling), which are also discussed in this thesis. In sum, this thesis makes a new contribution to corpus linguistic research on learner English and will have implications for the development of teaching practices for Malaysian learners of English.

Acknowledgements

In the name of God, the Most Gracious, the Most Merciful. Praise be to God for which this PhD journey has taken place, for the many ups and downs experienced, and by His leave, resulting in the accomplishment of this thesis.

My utmost respect and gratitude firstly goes to my supervisor, *Dr. Monika Bednarek*, who not only has guided me throughout this journey, but most importantly, has shown me what a great supervisor can be. I am also grateful to many educators whom I have known along the way, specifically *Dr. David Lee* (University of Wollongong) for acting as my associate supervisor in the early stages of my PhD, *Dr. Bronwen Dyson* (Postgraduate Writing Support) for teaching the Writing Workshop, and *Dr. Hajar Abdul Rahim* (Science University Malaysia, USM) for introducing me to and stimulating my interest in corpus linguistics in the first place. For providing the scholarship, which enabled me to pursue this PhD at the University of Sydney, I am grateful to the Ministry of Education in Malaysia and MARA University of Technology (UiTM), in particular the Academy of Language Studies. Thanks also go to *Bradley Smith* (PhD, Macquarie University) for his editing work on this thesis.

Throughout this journey, I learned that encouragement is definitely what you need and my heartfelt appreciation goes out to all my friends and family – near or far – for their never-ending support. To the new friends I made in Sydney, particularly those that are closest – *Habibah Ismail*, *Hasyimah Mohd Amin*, *Nurul Ikhmar Ibrahim*, and *Geeta Belraj*, your companionship is what I treasure most. To old and loyal friends, especially – *Jaja*, *Shima*, and *Leen*, thank you for always listening and cheering me on. My warmest love and thankfulness are for the people in my family. To both of my parents, *Dr. Joharry Othman* and *Azidah Othman* for being great role models and source of motivation. To my beautiful sisters, *Fatimah*, *Sara*, *Hajar* and loving brother, *Muhammad Khair* – thanks for being patient with me. And to my brother-in-law, *Azmi* and his kids, my in-laws, *daddy*, *ibu* and *Kak Sha*, for understanding.

Finally, I would like to express my sincerest gratefulness to my husband, *Muhammad Hafiiz Yahya*, for providing me with everything I needed and more, in finishing this PhD. This thesis has been a collective effort from the start and with that, I am truly happy and grateful for its completion.

Table of contents

List of figures

List of tables

Chapter 1: Introduction

1.1 Introduction

This thesis presents research about Malaysian learners of English, which uses a corpus linguistic approach to investigate recurring patterns in students' use of English in their argumentative essays. In Malaysia, the official language, Malay, is often spoken alongside other languages such as Mandarin and Tamil. Apart from the encouragement of bilingualism in Malaysia, the government also promotes English as a second language in the standard educational curriculum (Noor Abidah & Zaidah, 2008). One important part of the curriculum is the ability to write essays in English. This is mainly because essays are 'building blocks' for assessing English language skills (Schneer, 2014; Zhu, 2001); thus, part of the curriculum to enhance students' writing skills includes the teaching and learning of how to write argumentative essays. Furthermore, the evaluation of good writing skills is often assessed via argumentative essays. That is, argumentative writing is regarded as a common essay type in the English Language Teaching (ELT) classroom, and therefore is frequently assessed in all levels of examination in Malaysia.

However, argumentative or persuasive writing is a difficult mode of discourse for student writers, especially for second language (L2) users (Ferris, 1994; Schneer, 2014). This, as Ferris argues, is due to both "linguistic deficiencies and differing rhetorical patterns in the writers' first languages" (1994: p. 46). Furthermore, argumentative essays are primarily a social practice that requires the writer to construct a reasoned argument, usually involving "an awareness of audience [as well as] purpose and a mastery of necessary linguistic resources" (Morgan, 2011: p. 6). This means that interaction between the writer and reader is essential in writing argumentative essays. Linguistic analysis of learner language can be used to examine such interaction, alongside a wealth of other aspects of writing, including ways of constructing arguments. Although it is desirable for students to achieve skills in argumentation, Botley (2014) argues that it is particularly challenging to teach students these skills, given the complexity of arguments in discourse. Students, particularly at university level, are required to identify, produce and evaluate often complex reasoning in their studies. However, "it may not be enough to simply teach them how to write argumentative essays in a somewhat mechanistic and linear fashion, nor to identify and evaluate arguments using simple and canonical examples from textbooks" (Botley, 2014: p. 47).

With respect to the corpus approach to learner language, adopted in this thesis, Huat (2003) argues that language learning, specifically in writing, is best examined through contextualization of a pedagogic, or topic- and genre-based corpus. More specifically, Huat (2003: p. 48) argues that specialised corpora, which are based on recurrent topics and relevant genres, are potentially useful for exploring and investigating learners' writing in improving their language skills. This helps linguists in the process of investigating meaning and analysing linguistic data for the study of a target language, in this case, English. Hence, the present thesis intends to examine the linguistic patterns/features of Malaysian learner writing, by focusing on students' argumentative essays. Essentially, this thesis presents a corpus-driven, contrastive analysis of Malaysian learners' persuasive writing as compared against a comparable collection of native speaker writing, the latter of which acts as a point of reference rather than a norm (as explained in Chapter 3). In this chapter, I introduce the major motivations for the study, including a review of the language background and expanding field of corpus linguistics in Malaysia. I then situate the research within the theoretical framework of Contrastive Interlanguage Analysis (henceforth, CIA), and briefly describe the methodological contributions of corpus linguistics, highlighting the significance of keyword analyses that further point to new avenues for analysing discourse functions of lexical items. The benefits of such an approach for pedagogy are also briefly noted. Finally, I provide an outline of the thesis itself, including its main research questions, and an overview of the subsequent chapters.

1.2 Aims and rationale for the study

The incorporation of corpus methods into language research has been identified to have shown great value (Cheng, 2012; McEnery & Hardie, 2012). Among its many features are the generation of word frequency lists (alongside keywords lists, concordances and collocation, as explained in Chapter 3) and the ability to identify phraseological variation and promote statistical measures. With these techniques, the corpus linguistic approach allows for both quantitative and qualitative analysis. Another quality that makes corpus study more powerful and plausible than many other approaches is its availability to the public, and thus the ability of corpus studies to be investigated objectively from different angles and for different purposes. Since it is open to objective verification of results, the study of corpora, according to Leech

(1992), is a powerful methodology. Corpus linguistics has thus become well-established in a variety of fields, such as in discourse studies (Biber & Barbieri, 2007; Cheng et al., 2008), pragmatics (Aijmer, 1996), register (Biber et al., 1999; Scott & Tribble, 2006), genre analysis (Bednarek, 2006; Ooi, 2008), and – most relevant to this study – learner language (Flowerdew, 2009; Ishikawa, 2007; Paquot & Granger, 2012). Learner Corpus Research (LCR), as described by Botley and Dillah (2007: p. 77), "has developed into a well-defined field of research in recent years". LCR has paved the way for further research, and similar studies can be found elsewhere in the world. These studies are presented and discussed in Chapter 2.

In Malaysia, the study of language using corpus methods is continuously developing, particularly in the area of English language teaching and learning (Normazidah et al., 2012). This is probably due to the linguistic demands of using English in a rapidly globalising and modern society, as well as in attaining the country's vision to become a fully developed nation by the year 2020 (Zuraidah et al., 2010). While corpus linguistic studies are thus nothing new in Malaysia, contrastive corpus studies are limited (Siti Aeisha & Hajar, 2014). In a bibliographic analysis of corpus-related studies published between 1996 and 2012 in Malaysia, it was found that research has been focussed mainly on five areas: English use in Malaysia, Malaysian English learner language, Malaysian textbook content, Malay language description and lexicography, and corpora development (Siti Aeisha & Hajar, 2014: p. 19). Although there are a number of Malaysian corpus studies that employ the contrastive approach, some have only focussed on descriptive findings of particular groups of students, while others rely heavily on quantitative data (usually focussed on form/grammar) rather than on qualitative analysis. Most of these studies focus solely on the use or misuse of certain grammatical items in learner language, resulting in them being mostly descriptive (Mohamed Ismail et al., 2013; Nor Hafizah et al., 2013). Furthermore, interpretations of findings tend to generalise learners' language 'inaccuracies' rather than attributing them to other (external) possibilities, such as learners' multiple L1 background, genre of writing, and/or essay topics (Mukundan et al., 2013; Yunisrina, 2009). Individual lexical items are also overly emphasised in Malaysian corpus research, while analyses of phraseological patterns are scarce (Kamariah & Su'ad, 2011; Noorzan, 1998). More importantly, although the present scholarship of learner corpus studies in Malaysia has revealed significant insights into Malaysian's English language, more contrastive corpus method studies are anticipated in Malaysian LCR (Botley, 2010).

Therefore, LCR, underpinned by Contrastive Interlanguage Analysis (CIA), provides practical solutions in analysing data from the bottom-up, i.e. examining key features of the specific learner language that are extracted using corpora tools. Essentially, CIA is a methodological framework that enables two varieties of the same language to be compared, specifically native language vs. learner language ('Interlanguage' or IL) (Gilquin, 2001: p. 98). The present thesis aims to extend CIA within the scope of Malaysian LCR, comparing English written by native speakers of the language against written English used by Malaysian speakers. As will be pointed out in Chapter 3, many researchers have investigated learner language in writing, particularly via CIA, using corpus-driven methods. However, there remains no in-depth study on the investigation of Malaysian learner writing that includes contrastive analyses between comparable corpora (including reference language varieties), exploration of keywords analysis, as well as examination of discourse functions of salient items related to the genre and topic of essay writing.

The gap to be filled will be in exploring this further, through a study in which the description and evaluation of Malaysian learner English argumentative writing is compared to a comparable reference language variety, namely the Louvain Corpus of Native English Essays (LOCNESS). In fact, LOCNESS is rarely found to be compared with a Malaysian learner corpus.[1] In addition, findings of *both* salient individual lexical items *and* recurrent word combinations are significant to the description and evaluation of Malaysian learners' lexico-grammatical patterns in writing. In an attempt to adhere closely to the theoretical principles involved in conducting such type of research, detailed examination of frequency counts and statistical measures, along with innovative analyses of range and distribution, contribute to the existing knowledge of Malaysian learners' English, especially with regard to the demographic profile of Malaysian Corpus of Students' Argumentative Writing (MCSAW) writers, the genre of argumentative writing, and essay topics.

1.3 Background: The language situation in Malaysia

Before describing the theoretical framework of the present thesis, it is important to present a brief overview of background information on Malaysia and its language situation. Malaysia is a Southeast Asian country constituting the Malaysian Peninsula and parts of the island of

[1] Botley (2010) is an exception.

Borneo (Sabah and Sarawak). The country is multi-cultural: about half the population is ethnically Malay (50.1%),[2] with large minorities consisting of Malaysian Chinese (22.6%), Malaysian Indians (6.7%),[3] and various groups of indigenous people (1.8%). The constitution declares Islam as the state religion, while allowing freedom to practise other religion/beliefs. The government system is a constitutional monarchy, in which the head of state is the king, known as the *Yang di-Pertuan Agong*, while the head of government is the prime minister.

Given its multi-ethnic society, Malaysia is rich with diverse languages, ranging from the three main languages Malay, Mandarin and Tamil, to over a hundred types of indigenous languages such as the Iban language in Sarawak, and Dusun and Kadazan languages spoken in Sabah. There are also some 42 languages that are known to be endangered in Malaysia (as cited in http://www.endangeredlanguages.com/lang/country/Malaysia). Although Bahasa Malaysia (Malay in short)[4] is the official language of the country, English as a second language plays a part in many areas of communication, particularly in education, as noted above. In this thesis, I use *(Malaysian) learner English* to refer to the variety of English produced by learners of English in Malaysia. In contrast, *Malaysian English* refers to the local variety of English in Malaysia, which can be compared to other English varieties around the world like Singaporean English and Mandarin English (Hajar, 2006: p. 4; Imm, 2009: p. 451). Malaysian English is included in corpora such as the Malaysian sub-corpus of the International Corpus of English (ICE), which is further discussed in Section 2.2.1. In this thesis, I will only occasionally draw on Malaysian English where it seems relevant to the discussion of learner English.

The Malaysian education system and Malaysia in general have seen tremendous change, particularly since Malaysia attained its independence from Britain in 1957. Specifically, ELT in Malaysia was introduced by the British Government sometime in the early-nineteenth century. Since then, the use and importance of the English language has passed through many phases (Foo & Richards, 2004: p. 229). The Third Malaysia plan states that Bahasa Malaysia (Malay) is the basis for national integration and that English is taught as a second language (Saadiah, 2009; David, 2004). The education system is divided into preschool education, primary education, secondary education, post-secondary education, and tertiary education. Similar to many developed countries in the world, Malaysia adopts a system of 6+5+2 years of

[2] Accessed from www.livepopulation/malaysia.com on 8th of August 2016.
[3] Malaysian Chinese and Malaysian Indians would represent Malaysian-born Chinese and Indians whose earlier generations settled in the country during the British colonization era/period.
[4] Both Bahasa Malaysia and Malay will be used interchangeably to mean the official language in Malaysia.

formal education (i.e. primary, secondary, and post-secondary education/pre-university), in which English is a core subject and compulsory for all students (StudyMalaysia.com, 2015).

In realising English as a second language, Malaysia has experienced three major changes in language policy since 1994. Despite some criticisms from the Association of Malay Teachers, the government proceeded with the teaching of Mathematics and Science in English as one of their initiatives to promote Malaysia as an industrialised nation (David, 2004). Teachers who specialised in these respective fields were consequently retrained in an effort to enhance their proficiency and confidence in teaching the subjects in a language other than that which they were used to. Another major change, according to David (2004), is the increased number of private institutions since 1996 – from 50 to 650 colleges. Many of these institutions have twinning programs with foreign universities, and in turn, adopt English as the medium of instruction. "In 2000, … English was reintroduced as a subject in pre-university classes [and] [s]tudents who wish to enter local universities must sit for the Malaysian University English Test (MUET)" (David, 2004: p. 10).

Furthermore, the Malaysian education system follows two main sets of curricula: the New Primary Schools Curriculum (Kurikulum Baru Sekolah Rendah), implemented in 1983; and the Integrated Secondary Schools Curriculum (Kurikulum Bersepadu Sekolah Menengah), implemented in 1989; which have been revamped into, respectively, the Kurikulum Standard Sekolah Rendah (KSSR) and Kurikulum Standard Sekolah Menengah (KSSM)[5] (Saadiyah, 2009: p. 22). Saadiyah adds that,

> [t]he focus of the New Primary Schools Curriculum for [ELT] was the acquisition of the 3 R's namely basic skills of reading, writing and arithmetic. Moral and spiritual values were infused into the teaching of English in the Integrated Secondary Schools Curriculum through listening, speaking, reading and writing activities. Teachers were required to promote learners' intellectual development by posing questions that call for higher order thinking skills. Active participation from learners was also expected. The Integrated Secondary Schools Curriculum for English was a skill-based syllabus advocating Communicative Language Teaching (Kementerian Pendidikan Malaysia, 1989) and lessons integrated the four skills (Saadiyah, 2009: pp. 22-23).

[5] According to Education Director-General Tan Sri Dr Khair Mohamad Yusof, in a recent interview with the New Straits Times, the Curriculum Review, which was planned in the Malaysia Education Blueprint (2013-2025), was completed. He states that "The Education Ministry has completed the Curriculum Review for both primary and secondary schools, to be used in 2017 for all subjects". Accessed on 22nd August 2016: http://www.nst.com.my/news/2016/06/151751/revamped-school-curricula-next-year.

Since acquiring the writing goal of the 3Rs is important in attaining language competency, argumentative-type essays are continuously taught and assessed. According to Botley (2014: p. 45), "[a]rgumentative essay writing is a powerful pedagogical tool for developing and evaluating the ability of learners to construct sound and persuasive written arguments based on adequate logical support". He further adds that, in Malaysia, as elsewhere, the argumentative essay pattern taught in many programmes is more or less fixed in terms of its rhetorical structure. The essay firstly begins with an introduction and thesis statement, followed by at least three paragraphs containing topic sentences and a number of supporting statements, which in turn are summarised in the conclusion, including a restatement of the thesis (Botley, 2014: p. 46). Furthermore, topics of argumentative writing tasks usually concern contemporary social issues that encourage writers' demonstration of their general knowledge. A detailed explanation of the genre of argumentative writing is explored further in Chapter 3.

However, Nor Hafizah et al. (2013: p. 94) found that "Malaysian college students are unable to use interpersonal discourse in writing argumentative essays effectively". They argue that, besides facing difficulties in using interpersonal discourse, Malaysian learners also have problems using textual discourse effectively so as to produce a well-written argumentative essay. They conclude that difficulties arising among learners are most probably due to the limited range of students' vocabulary. In addition, they relate this problem to the lack of reading and writing skills of students, who tend to rely mostly on rote-memorisation. Students depend on this memorisation technique typically because of the over-riding concern for examination, in which researchers consider there is a mismatch between policy and practice in the Malaysian ELT curriculum (Normazidah et al., 2012: p. 42). The present thesis will further explore Malaysian learners' ability to produce argumentative essays, and identify areas for improvement that can be addressed in the teaching of English in the ESL (English as a Second Language) classroom in the future (see Chapter 7).

1.4 Theoretical and analytical framing

As noted above, this thesis is situated in Learner Corpus Research generally and Contrastive Interlanguage Analysis specifically. In LCR, employing the CIA framework is not only advantageous but widely popular (e.g. Gries and Deshors, 2014; Lee and Chen, 2009; Paquot and Granger, 2012). From a methodological perspective, the study of learner language through examining the Malaysian corpus (MCSAW) is best conducted via comparing it with a reference corpus (LOCNESS) rather than a norm. This is because LCR has often been criticised as comparing learners' performance against a native speaker norm (Granger, 2015); thus, the use of a reference language variety is seen to be valuable. Of most significance, and crucial to the present study, is Granger's (2015) revival of the CIA approach, which not only argues for the capabilities of analysing learner language via corpus methods, but particularly advocates the use of appropriate reference language variety for comparative purposes. This, as Granger (2015) highlights, is considered to be the best way for exploring learner language: emphasising comparable corpora, which, if conducted appropriately, are able to reveal to researchers the key traits of learner language relative to the reference language variety that is being compared. Thus, two major areas underpinning the present thesis - Learner Corpus Research (corpus linguistic research on learner language), and CIA – will be introduced in more detail in Chapter 2.

Briefly here, corpus linguistics is an empirical method for examining bodies of language known as corpora. By employing corpus software, such as WordSmith Tools (Scott, 2012), which is used in this thesis, it has become more feasible to examine linguistic items according to various types of analysis. Sophisticated corpus techniques are available to extract, calculate and reveal findings offering insights for linguists to understand aspects of language that were previously tedious or unfeasible to examine. In this thesis, such corpus techniques are used to explore both salient individual words as well as salient lexical bundles. This is in response to limitations of past research that focuses on individual and recurrent word combinations separately (Paquot & Granger, 2012).

Furthermore, a contrastive corpus-driven approach presents a range of methods that make investigating learner language more feasible than traditional types of analysing (e.g. Contrastive Analysis; Error Analysis). In Chapter 3, further discussion will be provided for the corpus-driven research, which adopts a bottom-up, inductive approach to language, identifying and interpreting frequently occurring items and the patterns in which they occur (e.g. Partington, 2004; Römer, 2004; Webb, 2010). Such an approach is considered to be strictly committed to "the integrity of the data as a whole" (Tognini-Bonelli, 2001: pp. 84-85), resulting in empirical findings. In the present research, the inductive, corpus-driven approach means taking keywords (statistically significant words) as a starting point, analysing the distribution of words across and within corpus files (range and dispersion), examining co-occurring words (collocates), and further examining concordance lines for qualitative analysis. It must be noted that the texts from which keywords are extracted may contain errors, and in interpreting results, I relied on my own understanding of these error-laden samples as consultation with external experts was not possible.

In addition to being corpus-driven, the approach adopted in this research is contrastive: that is, it involves the process of comparing and contrasting two argumentative-type written corpora, MCSAW and LOCNESS. Chapter 3 further elaborates how the two corpora compared in this thesis are highly relevant for comparison purposes. For Malaysian learners and educators alike, the investigation of learner language in MCSAW is valuable to the understanding of Malaysian learner language, as well as to provide insight into students' proficiency in writing English, specifically in the area of argumentative essay writing. By comparing two sets of written argumentative texts, we can recognise styles of learner writing that may be indicative of the genre of argumentation or indicative of their writing tendencies as a whole. In addition, comparison with a reference language variety offers insights into the differences between the novice writers of MCSAW and LOCNESS. Novice writers are authors of "unpublished pieces of writing that have been written in educational or training settings" (Scott & Tribble, 2006: p. 133), rather than authors of "expert texts […] that have been published" (Römer, 2009: p. 149). Expert writers tend to have "better experience and knowledge of the field and/or greater facility with the language" (Lee and Swales, 2006: p. 68). Rather than comparing expert with novice writing, this thesis compares the novice writing produced by two different groups of writers. In turn, differences or similarities between these two groups can be identified. Other benefits of the contrastive approach include highlighting the effects of ELT in Malaysian classrooms, and alternative ways to enhance better performance in attaining language competence and proficiency overall. Addressing the lack of contrastive corpus-driven research in investigating Malaysian learner language, the present study contributes to the scholarship on corpus linguistics and the practicality of using this methodology in explaining and interpreting learner language, specifically with regard to Malaysian learners of MCSAW.

1.5 Research questions

The present thesis thus sets out to answer four research questions:

1) What are the most salient linguistic items found in the Malaysian learner corpus compared to those in the reference corpus?

2) How are the items used similarly or differently in the two corpora (including their collocations)?
3) What are the most overused types of lexical bundles found in MCSAW?
4) How do these bundles function in Malaysian learner argumentative writing?

The present study aims to contribute to the existing body of knowledge within the field of contrastive corpus-driven studies in Malaysia and current ELT instruction, particularly through examining Malaysian learners' argumentative writing. Following CIA as the theoretical framework of analysis, this study focuses on learners' key linguistic items in their argumentative texts, using corpus tools. By investigating both individual and lexical bundles, as outlined in the research questions above, I expand the use of corpus linguistics for more than one linguistic phenomenon/area, for reliable interpretation of empirical data (Section 3.1). Furthermore, to ensure the comparability and effectiveness of the contrastive approach, the thesis highlights that the target corpus, MCSAW, presents some issues, mainly because it only includes two essay topics, revealing much repetition between essays, as the use of the learner corpus in this thesis will reveal (Section 3.2). Thus, the present study is contrastive and corpus-driven, revealing two important observations: CIA is (again) ideally effective when comparability issues are addressed; while the bottom-up approach further strengthens the validity and reliability of the research findings. Nevertheless, findings show that, despite the salient features of learner writing compared to the reference language variety, essay topics play a significant role in learners' written tasks (Chapters 4, 5 and 6).

1.6 Structure/overview of thesis

This chapter has provided the background to the study, and the study's objectives, and placed these within the context of Malaysian argumentative essay writing. In so doing, it has provided brief introductions to some relevant terms. Other terms and methodological issues are outlined in Chapter 3. The remainder of this thesis consists of a further six chapters.

Chapter 2 provides a theoretical and analytical framework for the study by reviewing current literature on corpus linguistics and LCR, and on the Malaysian context of learner corpus studies in particular. Literature pertaining to methodology for LCR, specifically for the CIA approach, is also reviewed in Chapter 2.

Chapter 3 explains the methods that have been used for conducting the research and for the analysis of the data used in this study. It describes the methodology and data (corpora), the operational procedures used for analysing data, the selection of data for analysis, and the qualitative methods of analysing concordance lines.

Chapters 4 and **5** present the results of individual keywords analysis based on a selection of the most salient features in Malaysian learner writing as compared to the reference corpus, and provide discussion of these findings according to past research on modality and personal pronoun use. Chapter 4 reports results of the prevalent use of the modal verb *can* in learner writing, and gives some insights into learners' various uses of this modal's meanings. Chapter 5 then reports results of the salient use of the personal plural pronoun *we* in learner writing, and provides insights into its discourse functions.

Chapter 6 proceeds with results of the analysis of key lexical bundles that are unusually more frequent in the learner corpus in contrast to the reference language variety. It gives some insights into the use of lexical bundles, which are regarded as chunks in constructing language.

Chapter 7 summarises the research findings and contributions of the study, accounting for the study's limitations, proposing directions for future research, and addressing the ways in which these findings might inform future curricula that efficiently and effectively empower Malaysian learners in their pursuit of academic literacy.

Chapter 2: Learner Corpus Research

2.1 Introduction

Corpus linguistics has been shown to have immense value for the study of language, specifically given the use of naturally attested data (corpora). Corpus studies are revolutionising the study of learner language, given the investigation of frequencies, functions and contexts of words in learner language (Biber et al., 1994; Staples & Reppen, 2016). Moreover, corpora are regularly used for validating hypotheses (Aarts, 2000). The literature on corpus linguistics is extensive (McEnery & Hardie, 2012); thus, this chapter focuses solely on corpus linguistic research on learner language. Learner language is here defined as data (either spoken or written) derived from foreign or second language speakers of a particular language.

As introduced in the previous chapter, the present thesis aims to conduct a contrastive investigation of Malaysian learners' argumentative writing against a comparable native-speaking reference language variety, using corpus methods. In this chapter, a survey of learner corpora is firstly presented, followed by a brief summary of other types of corpora presented (Section 2.2). Benefits of using corpora in language studies are also discussed. A review of the literature highlights two dominant approaches in learner corpus research, Contrastive Analysis and Contrastive Interlanguage Analysis, of which the latter is seen to be more prevalent in the field (Section 2.3). Criticisms of earlier methods are also mentioned and evaluated, clarifying the rationale for the recent approach to be taken in the present thesis. Finally, Section 2.4 discusses how corpora have been used in linguistic analysis related to the investigation of learner language, the connection between novice writing and spoken features in writing, and how the present thesis can contribute to existing scholarship.

2.2 What is a corpus? – Corpus linguistics and the use of corpora

Corpus linguistics (henceforth, CL) can be loosely described as an approach to studying language using naturally-occurring data. While there are many definitions of CL (Adolphs & Lin, 2011; Cheng, 2012; McEnery & Hardie, 2012; Sinclair, 2004b), for the purpose of this thesis, CL will be regarded as the process of analysing and theorising language that can be done by examining amounts of real, empirical data, alongside sophisticated computerised software tools (Lee, 2008: p. 87). In other words, this gives linguists the power to search, process, and analyse language without the difficulties of compiling, counting and describing language manually. This section begins with a survey of corpora, specifically on learner corpora. Then, emphasis is given to corpus studies that are focussed on learner corpus research, including studies that are situated particularly in Malaysia.

2.2.1 Learner corpora

A corpus is a body of language representative of a particular variety of language or genre – collected and stored – mostly in electronic form, which can be used for analysis using concordance software (Baker, 2006: p. 25). A survey of corpora developed over the years shows a positive growth both in size as well as types of corpora built (Lee, 2010).[6] The majority of corpora comprise English written texts (e.g. Corpus of Contemporary American English; Longman Written American Corpus), while spoken corpora of English are significantly less numerous (Cambridge & Nottingham Corpus of Discourse in English; Michigan Corpus of Academic Spoken English). This is mainly because collecting, analysing and transcribing oral data are more challenging than compiling texts that are written. For diachronic research purposes, historical corpora can be compared with contemporary ones in order to investigate language change over time. In addition, specialised corpora offer examination of specific dialects, genres, and registers. While most of the general corpora are in English and produced by speakers of English, there is a growing development of learner corpora (i.e. texts by learners of English), some of which are in other languages, which are described next.

Learner corpora, i.e. "electronic collections of writing or speech produced by foreign or second language learners" (Gilquin & Granger, 2015: p. 1), are described as a relatively new addition to the wide range of existing corpus types (Nesselhauf, 2004). Over the past years, Granger and Dumont (2012) have made a comprehensive list of learner corpora around the world, while inviting others to contribute to this on-going list (http://www.uclouvain.be/en-cecl-lcworld.html). Learner corpora are collected following a strict design criterion (Granger, 2008: p. 344). Some of the criteria used for the compilation of learner corpora include language, medium, text type(s), level(s) of learners, L1 or first language(s) of learners, and task setting (Nesselhauf, 2004: p. 130). The basis for compiling corpora according to these criteria is to

[6] Also see http://www.uow.edu.au/~dlee/corpora.htm, 'Corpora, Collections, Data Archives'.

"control the wide range of variables that affect learner language, both learner variables (age, proficiency level, mother tongue background, etc.) and task variables (field, genre, topic, etc.)" (Gilquin et al., 2007: p. 322). In addition, careful design criteria avoid a biased selection of data and allow for comparative studies (Gilquin et al., 2007: p. 322). Meanwhile, the main purpose in compiling a learner corpus is to gather objective data that can aid in the process of describing learner language, particularly use of language by learners in actual production (Gilquin & Granger, 2015; Granger, 1998).

Similar to many other types of corpora, English is the target language in most learner corpora (e.g. The Advanced Learner English Corpus/ALEC; The Chinese Academic Written English corpus/CAWE). Learner corpora that focus on languages other than English, such as Spanish, Italian, and German, are still few, but have been shown to contribute to the amount and variety of learner data besides those in the English language (Gilquin & Granger, 2015). Other types of learner corpora comprise more than one language, in that they are multilingual, such as the corpus PARallèle Oral en Langue Etrangère (PAROLE), which consists of texts in English, French, and Italian. Also noteworthy are Chinese learner corpora, such as The Jinan Chinese Learner Corpus (JCLC).

Learner corpora, which can include language produced by learners of different origins and different proficiency levels, can be categorised into different types, including general or specific, written or spoken, synchronic or longitudinal, and mono-L1 or multi-L1 data (Gilquin & Granger, 2015: p. 418). In the following discussion, three major learner corpora are mentioned in more detail.

A pioneer learner corpus is *The International Corpus of Learner English* (ICLE), as it was "the first learner corpus created in an academic setting" (Pravec, 2002: p. 83). ICLE comprises argumentative essays written by advanced learners of English (i.e. university students of English in their third or fourth year of study) from various native language backgrounds, namely Bulgarian, Chinese, Czech, Dutch, Finnish, French, German, Italian, Japanese, Norwegian, Polish, Russian, Spanish, Swedish, Tswana, and Turkish. The corpus, which was launched in 1990 by Sylviane Granger, is highly homogeneous (as all collaborative universities have adopted the same corpus collection guidelines)[7] and is continuously being developed at the Universite Catholique de Louvain (UCL) in Belgium. The *Louvain International Database*

[7] Corpus collection guidelines can be viewed at http://www.uclouvain.be/en-317607.html.

of Spoken English Interlanguage (LINDSEI) is the spoken counterpart to ICLE, containing oral/speech data produced by advanced learners of English from several mother-tongue backgrounds. Other learner corpora consist of both written and spoken texts, such as *The International Corpus Network of Asian Learners of English* (ICNALE), and The LONGDALE project: *LONGitudinal DAtabase of Learner English*. However, one caveat lies in the availability of learner corpora that may require one to retrieve passwords or obtain permission from the corpus developers (e.g. *The University of Toronto Romance Phonetics Database,* RPD).

Another look into the existing scholarship identifies a rise in studies pertaining to Asian learners. ICNALE (the *International Corpus Network of Asian Learners of English*), which was compiled by Ishikawa (2011), consists of student essays from a number of Asian countries, namely China, Indonesia, Korea, Japan, Hong Kong, Pakistan, The Philippines, Singapore, Taiwan and Thailand. Corpus studies in the Asian region, especially, have benefited from ICNALE, and have shown many useful insights into learners' development in learning a language (e.g. Hu & Li, 2015; Ishikawa, 2014). Some of the advantages include the benefits of using multi-L1 corpora,[8] the practical use of an online corpus tool (WordSketch), and adapting the contrastive interlanguage analysis, to name a few.

Although the majority of learner corpus studies are based on raw data, there is an increasing number of studies that make use of annotated data, usually in the form of part-of-speech (POS) tagged or error-tagged data. These include studies such as Granger (2003) and Bestgen and Granger (2014). However, annotation of learner data has been argued to be problematic, as POS-taggers were not found to perform as well on learner texts as on native corpus data (Gilquin & Granger, 2015: p. 2). For this reason, annotated learner data will not be used in the present study. In summary, learner corpus data offer a number of significant advantages:

> the corpora are usually quite large and therefore give researchers a much wider empirical basis than has ever been available before; they can be submitted to a wide range of automated methods and tools which make it possible to quantify learner data, to enrich them with a wide range of linguistic annotations (e.g. morpho-syntactic tagging, discourse tagging, and error tagging) and to manipulate them in various ways in order to uncover their distinctive lexico-grammatical and stylistic signatures (Gilquin et al., 2007: p. 322).

[8] Refer to Altenberg and Granger (2001) with regard to multilingual corpora and cross-linguistic studies.

In Malaysia, a survey of CL studies shows that there was a rise in the amount of published research using corpus methods between the years 1996 to 2012 (Siti Aeisha & Hajar, 2014). Although corpus study was firstly introduced in the creation of a Malay language corpus in the early 1980s, Malaysian corpus research in English is shown to have begun in the 1990s, and is continuously growing (Hajar, 2014). Various types of corpora have been produced within the Malaysian corpus research scene, mainly English language learner corpora such as the English of Malaysian School Students (EMAS) corpus (Arshad et al., 2002), Malaysian Corpus of Learner English (MACLE) (Knowles & Zuraidah, 2004), and Corpus Archive of Learner English Sabah-Sarawak (CALES) (Botley et al., 2005). Others consist of genre-specific learner corpora such as the Engineering Lecture Corpus (ELC) and the Business and Management English Language Learner Corpus (BMELC), as well as the development of English pedagogic corpora (Mukundan & Menon, 2007). Similarly, the corpus used in the present study (MCSAW) is regarded as a genre-specific learner corpus, as mentioned in the previous chapter. This corpus, i.e. MCSAW, will be introduced in Chapter 3. In short, learner corpora are continuously being collected and designed for a variety of purposes. Given that English is mainly the language being investigated among these learner corpora, it is becoming more noticeable that there is an increasing population of non-native speakers learning English in the world and, hence, studies on learner language using corpora are worth being explored. The next section will introduce such research.

2.2.2 Corpus linguistics in language studies: Learner Corpus Research

According to Granger (1998), learner corpus research (henceforth, LCR) is interdisciplinary; studies within this field explore many facets of language including foreign language teaching, corpus linguistics, natural language processing, and second language acquisition. In her 1998 collection of papers, Granger reported that most of the research was done on comparisons between native speaker English and learner English. This has been an on-going trend in LCR. For many linguists, such as Nesselhauf (2004: p. 126), the best way to find out learners’ typical difficulties with a certain language is “to analyse the language produced by a certain group of learners and compare it with the language produced by native speakers”. Hunston (2002a), in agreement, highlights that essentially there can be two types of comparison: “between corpora produced by different sets of learners, and between corpora produced by learners and those produced by native or expert speakers” (Hunston, 2002a: p. 206).

Granger’s (1998) book was among the earliest collections of CL studies in learner language, and it has significantly encouraged more interest and awareness in the area. Essentially, a look into these collected works (Granger, 1998) reveals three general aspects of learner corpora, namely that learners: have a tendency to use a smaller range of vocabulary items; over-use certain vocabulary items of high generality; and use more spoken features of language in their writing (Hunston, 2002a: p. 207). While these findings have been insightful in LCR, Hunston argues that “more investigation is needed before advice to learners can be given” (2002a: p. 208).

LCR has further developed in recent years (Gilquin & Granger, 2015; Ishikawa, 2014). This has involved an increase of the number of corpora collections beyond an exclusive English focus to one on a wider range of other languages (e.g. Götz & Schilk, 2011), the development of annotation and design of error-tagging systems (e.g. Thewissen, 2013), a much wider spectrum of linguistic analysis, i.e. phraseology (e.g. Bestgen & Granger, 2014), a growing integration of second language acquisition theory (e.g. Ädel & Erman, 2012), and vast applications of resources (e.g. Miller et al., 2016). There has also been an evolving analytical methodology, from the traditional Contrastive Analysis to the Contrastive Interlanguage Analysis (e.g. Lee & Chen, 2009), as discussed in Sections 2.3.2 and 2.3.3 below.

While LCR studies in general have seen much progress in recent years, the scholarship on learner corpus studies in Malaysia is still developing. As Hajar (2014: p. 7) notes, LCR in Malaysia could benefit more from the production of spoken Malay and English corpora, the development of Malay learner corpora, and of multimodal corpora.

In general, there is a strong consensus that LCR studies focus more on the description of learner language than its interpretation (Ellis & Barkhuizen, 2005). Granger et al. (2013) argue that, while there is positive progress on learner corpora in the CL scene, much is to be done to minimise the gap between LCR and second language acquisition (henceforth, SLA).[9] Although both fields (SLA and LCR) investigate learner language, SLA studies focus on competence, whereas LCR studies focus on performance (Gilquin & Granger, 2015: p. 418). In addition, the former group uses more manual and traditional SLA methods of analysis, suitable for the investigation of a small number of individual learners, while the latter group applies automated tools and techniques of corpus linguistics (Gilquin & Granger, 2015: p. 418). Unlike the more

[9] Similar arguments are raised with regard to CL and theoretical linguistics in Gries (2010) and Barlow (2011).

experimental data types often used in SLA, where learners are made to produce a particular form (e.g. elicited-type tasks), the focus in learner corpus data is on message conveyance and the possibility for learners to use their own wording (Gilquin & Granger, 2015: p. 1). More specifically, this means that investigating learner language via learner corpus data ensures that naturally-occurring data[10] are examined. Granger and her team are strong advocates for greater attention to theory or SLA-led research, in which replication of SLA studies can be carried out using corpora, and a more systematic integration of learner corpus-informed insights into pedagogical and natural language processing tools (see Granger, 2012; and Barlow 2005).

Sophisticated automatic processing tools used in LCR also give power to a combination of both quantitative and qualitative analyses, which are both equally important for theorising purposes. For example, many corpus studies start off with descriptive statistics, using corpus software tools such as AntConc (Anthony, 2012) and WordSmith (Scott, 2012) to extract lists of highly-frequent words, for example, followed by use of concordancing that allows linguists to make further interpretation of patterns of language in context. Functional analyses are also applicable, in which discourse functions of particular items are analysed to discern their use in specific contexts. In addition, CL promotes total accountability, in which a description of all the data in a respective corpus can be easily reported (McEnery & Hardie, 2012). This makes it possible to reveal new facts about learner language (e.g. patterns of word combinations, overuse/underuse of certain words, etc.), often due to the 'bottom-up'[11] (combined with top-down) processing of data. This is facilitated by the incorporation of large collections of texts as well as computer software tools, which are discussed in Chapter 3.

Essentially, generalisations can be made with regard to certain aspects of learner language, namely through the investigation of frequency, collocations and keywords analyses that are offered by CL studies. The next part of this chapter describes these concepts in relation to corpus studies on learner language, which will underpin the theoretical framework of the overall thesis.

[10] Following Nesselhauf (2004: p. 128), "what comes closest to naturally occurring texts in its strict sense are texts that are produced for pedagogical reasons and texts that are elicited for the corpus but that use procedures exerting very little control".

[11] See Cheng (2012: pp. 30, 176, 187-9, and 211).

2.3 Approaches to LCR Studies

There are various ways in which corpus linguists have looked at learner language, especially with regard to writing. Generally, LCR can be categorised as studies that are ‘corpus-based’ or ‘corpus-driven’ (McEnery & Hardie, 2012). As introduced in Chapter 1, the difference lies in the importance of the initial assumptions that govern a particular research and the role that the data play in the analysis. In the present chapter, the term ‘corpus studies’ will be used as an umbrella term to relate to both types, corpus-based and corpus-driven studies. However, for the purpose of this review, it is more useful to classify the scholarship in LCR into two main categories: non-contrastive and contrastive studies. The following part of this chapter will review the relevant scholarship of these categories, and how, recently, there has been an increasing popularity in comparative studies of the investigation of learner writing.

2.3.1 Non-contrastive studies

One of the typical ways to investigate learner writing in CL is to conduct an error analysis (EA).[12] EA, for the most part, has emphasised only the scrutiny of errors, and while learning from these errors can give insight into learners’ SLA, it can be seen as merely descriptive. According to Granger (1998), early learner corpora “were not really exploited as corpora in their own right, but merely served as depositories of errors, only to be discarded after the relevant errors have been extracted from them” (p. 6). In turn, there was not much development of the investigation of learner language, given these isolated works of EA. In addition, corpus studies that adopt the error analysis method identify errors according to the EA framework, and usually compare errors to dictionaries as well as the British National Corpus (BNC) for acceptability. Ang et al. (2011), for example, identify errors produced by Malaysian writers in a corpus of secondary-level school students, and hypothesise based on these errors that Malaysian learner language is systematic and can be influenced by both interlingual and intralingual factors. Evaluations of errors in this study were based on the Oxford Collocations Dictionary and BNC, which raises some criticisms in explaining patterns of language based on a corpus, as well as comparing learner writing with a general corpus of native speaker norm

[12] EA has developed into a new discipline, computer-aided error analysis (CEA), more rigorous methodologically, and therefore more apt to result in ‘learner-aware’ and efficient pedagogical tools (Gilquin, 2001, p. 97).

(Granger, 2015). Granger explains that the continuous reference to native speaker language in many LCR has created "recognition of the existence of one single monolithic norm in L2 studies" (2015: p. 15), and therefore, suggests that comparisons of learner data can be set with a large number of different reference points, i.e. reference language varieties. Another study that used dictionaries is Nesselhauf (2003), who conducted a corpus study on German learners of English and evaluated their language acceptability by reference to English dictionaries, namely the Oxford Advanced Learner's Dictionary and the Collins COBUILD English Dictionary. Apart from identifying the type of mistakes that the learners made when producing collocations, she also presented some of the reasons that might influence these mistakes, including the learners' first language (L1) backgrounds. Although findings from studies such as the above emphasise errors, they show some fruitful insights in relation to frequent collocation mistakes made by learners and how errors are likely to have been influenced by learners' L1s.

Studies that adopt the EA method without using a reference corpus have a tendency to evaluate errors solely by intuition, and could therefore become problematic (e.g. Darina et al., 2013), "since native speaker intuitions are not a reliable source of evidence" (Stubbs, 1995: p. 24). While errors play a significant role in the process of acquiring another language, errors are argued to signal a marked feature of a particular language variant (Lee & Chen, 2009). Lee and Chen contend that learners' writing may not necessarily be ungrammatical. In fact, this may make it more interesting to explore in describing learners' interlanguage (IL), which is described as "the knowledge of the [target] language in the speaker's mind" (Cook, 2014: p. 190). The argument is that investigations of a learner's IL do not only rest on the examination of errors learners make when compared to a standard set of guidelines (i.e. EA framework), or based on pure intuition. On the contrary, learner corpora can best be used to examine learners' IL processes by comparing significant frequencies and identifying idiosyncratic language patterns from a comparable set of reference corpora, i.e. a reference language variety (Granger, 2015).

While the above-mentioned studies have described learner language to a certain extent as being error-prone or influenced by the learners' mother-tongues, many studies have compared learner corpora with a reference or native speaker corpora in order to make wider generalisations. In investigating his Bruneian students' writing, Crompton (2005) investigated patterns of errors for the word *where*, which he found were problematic, and used two native

English speaker corpora (i.e., Longman Corpus of Spoken and Written English, LSWE; and the BNC) to make comparative analyses. Crompton states that it has become more feasible to compile small corpora of students' writing and make further analyses when compared to standard usage of English. In contrast to non-contrastive methods, as found in many EA studies, "learner corpora are often best used in combination with native speaker corpora" (Nesselhauf, 2004: p. 126). For this reason, the investigation of Malaysian learner's writing is compared against a native speaker corpus in the present research, for distinctive language patterns not visible when conducting a non-contrastive study. As the next part of the chapter will show, contrastive corpus studies are more promising in studies of learner writing.

2.3.2 Contrastive studies

As mentioned, CL has enabled comparisons between native speakers of English and learners of English to be made (Granger, 1998; Gilquin, 2015; etc.).[13] Most of these studies can be categorized within the EFL (English as a Foreign Language) context, specifically among French (e.g. Paquot, 2013; Thewissen, 2013), Spanish (Luzón, 2009; Martinez-Garcia & Wulff, 2012), Swedish and Finnish (Ädel & Erman, 2012; Peromingo, 2012), German (Rankin, 2012; Römer, 2009) and Chinese (Fawcett, 2013; Lee & Chen, 2009) learners. Based on these studies, linguists are highly interested in the non-nativeness and idiomatic features of expressions in learner writing, as well as how findings from their studies can benefit classroom pedagogy. Contrary to EA, that highlight errors and misuse of lexical items, linguists have explored learners' tendency to over- or under-use certain words/phrases. They have identified problematic areas such as over-using words/phrases, involving high-frequency common words, and confusing academic expressions. Based on these studies, it can be summarised that features of learners' writing exhibit a "still-developing interlanguage system" (Lee & Chen, 2009: p. 292).

[13] Translation studies is a related area of CL research. Studies that conduct these so-called 'cross-linguistic analysis' normally use a large English corpus (e.g. BNC, FLOB, and ICE-GB) as reference (e.g. Berber-Sardinha, 2000; R. Xiao & McEnery, 2006). Given the corpus software and tools that are able to work with various languages and the process of annotating, complex languages such as Chinese can be investigated easily with regard to their collocations and near-synonyms (Xiao & McEnery, 2006). In addition, evidence of semantic prosodies can also be identified and revealed by the exploration of two or more corpora (De Clerck et al., 2011; Zhang, 2009). Semantic prosodies, which involve the connotation conveyed by the regular co-occurrence of lexical items (Hunston, 2007), and are part of Sinclair's description of pragmatics expression (Sinclair, 1996), is a rather complicated aspect of linguistic investigation but is feasible through corpus methods and fruitful to explore in LCR studies (e.g. Oster, 2010; Partington, 2011).

While a majority of contrastive studies are taken from an EFL context, there are a number of studies that specifically investigate learner corpora in terms of ESL (English as a Second Language) learners (e.g. Laporte, 2012; Nesselhauf, 2009). It should be borne in mind that, for the purpose of the present thesis, the distinction between ESL and EFL learners is made based on how the target language (i.e. English) is used in the respective ESL/EFL learners' settings. According to Gotz and Schilk (2011), "[w]hile in the ESL-context, English is used for a variety of international as well as intranational functions, in EFL-communities, English is mainly used for international purposes or in restricted institutions" (p. 80). In the Malaysian context, learners are considered as ESL learners, since English is used as a medium of instruction in schools and universities (as described in Chapter 1). Owing to this as a potential factor in learners' language production, findings in the present thesis will be interpreted with this in mind. Overall, findings from these studies reveal that EFL and ESL learners share a number of features in terms of over-, under- and misuse of certain expressions of English (see further, Section 2.4).

As can be seen, the scholarship on corpus contrastive studies is vast in exploring learner language use. Contrastive studies enable linguists to examine what is particularly difficult for certain groups of learners when compared to native speaker corpora. More specifically, Hunston (2002a) highlights two major advantages over other methods of examining learner language use. She states that the basis of assessment is entirely explicit and realistic, where "learner language is compared with, and if necessary measured against, a standard that is clearly identified by the corpus chosen [and] that what the learners do is compared with what native/expert speakers actually do rather than what reference books say they do" (Hunston, 2002a: p. 212). This is also further discussed in Granger (2015), where the use of a particular reference language variety is carefully thought out in order for comparable results to be explained. In the present thesis, the reference corpus is chosen specifically in accordance with the novice writer in argumentative writing in mind. Furthermore, corpus analyses have been shown to provide more empirical findings through the process of counting, measuring of distribution, and significance testing. In this thesis, items are not only measured by how frequently they appear, but also how widely distributed they occur or how many times they occur in separate texts (using range).

Many contrastive studies have demonstrated the contrastive interlanguage analysis (henceforth, CIA) approach. In other words, comparative analyses were carried out between one or more languages or interlanguages (e.g. Dam-Jensen & Zethsen, 2008; Smith & Nordquist, 2012). Section 2.3.3 will, therefore, describe CIA in more depth, and review how studies have been conducted within this framework of analysis.

2.3.3 Contrastive Interlanguage Analysis (CIA)

Granger (1998) introduces an extension of contrastive analysis that observes what non-native and native speakers of a language produce in comparable situations. This approach is called Contrastive Interlanguage Analysis, i.e. CIA.[14] In general, interlanguage, or IL, can be described as the process by which learners reach the near-nativeness of a target language. More specifically, Selinker (2014: p. 223) defines it as a "linguistic/cognitive space that exists between the native language and the language that one is learning. [Thus,] [i]nterlanguages are non-native languages which are created and spoken whenever there is language contact".

There are two primary types of comparison in CIA: comparison of native language and interlanguage (IL); and comparison of different interlanguages (ILs). The first comparison "aim[s] to uncover the features of non-nativeness of learner language" (Granger, 1998: p. 13), specifically the over/underuse of some features of the language, requiring a control corpus of a native language. In both types of study, a specific learner group is examined against a native speaker corpus to compare frequencies and further discrepancies using corpus methods. For instance, Laufer and Waldman (2011: p. 665) found different types of English collocational errors for different levels of Hebrew learners, and they observed that these errors were partly due to the L1 influence. Similarly, in his study of causal links between German and British English writing, Lorenz (1999: p. 59) found that German learners produce substantial overuse of *because* and, interestingly, that German learners also showed a relatively lower rate of use of causal adverbs (e.g. *so, therefore, then, thus, hence, consequently, accordingly*). Consequently, the present research employs this type of CIA method, comparing the Malaysian group of learner argumentative writing against a similar text-type produced by novice native-speaking writers. Ways in which the two groups are compared will be presented in Chapter 3. This, in turn, will provide various meaningful insights into learners' strategies and preferences for the English language.

[14] It is now suggested that the approach of Contrastive Analysis can be combined with that of Contrastive Interlanguage Analysis (CIA), in what has been called the Integrated Contrastive Model (Granger, 1996; Gilquin, 2001).

The second most common comparative analysis is between two interlanguages, where studies compare ILs of the same language or of different languages (e.g. Martinez-Garcia & Wulff, 2012; Rankin, 2012). The aim is to investigate varieties of English in terms of various factors such as age, proficiency level, L1 background, task type, learning setting, and medium, among other things (Granger, 1998). These varieties of English (i.e. produced by Spanish, German, Dutch and French learners) can also be compared to a native speaker corpus. Given this, a comparison is not only examined in terms of the non-nativeness of each variety but also how different each variety is when compared to each other. Among their findings is that German and Spanish ESL learners share certain overgeneralisation tendencies, including the overuse of phrasal verbs and verbs in the gerundial construction (e.g. *continue, go on, keep on, end up, prefer*), but that German learners produce more native-like use of verbs in both gerundial and infinitival construction overall compared to their Spanish counterparts (Martinez-Garcia & Wulff, 2012: p. 240). Rankin (2012) discovered that German, Dutch and French learners appear to have difficulties in structuring their language regardless of grammar knowledge. The study further highlights that transfer of verb patterns in all the learners' writing may be the result of transferring patterns from the L1 rather than difficulty in the target language construction (Rankin, 2012: p. 155). While it is interesting to compare different types of learner languages in Malaysia, more available corpora are needed for this type of CIA research.

2.4 Investigating Linguistic Phenomena in Learner Corpus Linguistics

Given the rise of more research in corpus linguistics, LCR (specifically CIA) has undertaken various types of research on learner language, mostly on words (lexis), the continuity of lexis and grammar, as well as on discourse (Gilquin & Granger, 2015). The availability of learner corpora containing non-native writing makes it possible to carry out comparisons of specific linguistic features (Gabrielatos & McEnery, 2005). The next section describes how various linguistic phenomena are explored in learner corpus studies.

2.4.1 Word- and category-based: Lexis and Grammar

As can be seen in the previous sections, LCR studies can be discussed generally in terms of non-contrastive and contrastive studies. More specifically, many types of these EFL and ESL corpora studies can be discussed further as regards the types of linguistic features that are being investigated and how they are investigated, namely looking at lexis and grammar, multi-word combinations and phraseology. To start, this section will discuss learner corpus studies that have investigated lexis and grammar with respect to word- and category-based analysis (i.e. exploring single-word forms, or particular categories such as modality), as described by Hunston (2002b).

Almost all corpus studies start with examining the frequency and/or distribution of certain words or phrases in a text or collection of texts. Given the various uses and meanings of words in English, a lemma is often examined when a general inspection is to be made of a particular lexical item. This, i.e. lemma, means that the different inflectional forms, for example, *use, uses, used, using,* are merged (e.g. De Cock & Granger, 2004; Xiao & McEnery, 2006). While it is more reasonable to apply annotation techniques for investigations of lemmas or word forms, as mentioned earlier in Section 2.2, the present thesis does not delve into this matter, given the argument that using POS-taggers is challenging with learner corpus texts. Nevertheless, the advantages of examining word frequencies reveal interesting findings, in regard to what learners usually produce in their speech or writing. For example, Cobb (2003)[15] investigated all the words in the Quebec learner corpus in order of frequency, and found that learners over-use (almost 90%) common words that are within the 0 – 1000 frequency range (Cobb, 2003: p. 403). Consequently, studies that have investigated high-frequency words/phrases (e.g. Byrd & Coxhead, 2010; Coxhead, 2012) are interesting to refer to when it comes to the description of learner language findings. Other frequency-based LCR include Laporte (2012), and Salazar and Verdaguer (2009).

Linguists have also identified words that occur unusually more frequently when compared to a reference corpus, for their initial analysis. From this, linguists are interested in investigating single words and sometimes the recurrent words co-occurring around them (collocates), which could reveal possible insights into learners' word combinations in their language learning. One important study to mention involves the use of keyword analysis as the start of further linguistic analysis (Lee & Chen, 2009). In their study, Lee and Chen employed this "corpus-comparison technique" (ibid: p. 152), in which a list of keywords[16] is extracted automatically via corpus

[15] His study was an attempt to replicate Ringbom's (1998) study on vocabulary frequencies in advanced learner English.

[16] A keyword is generally a word that appears unusually frequent or infrequent in a target corpus, in comparison to its frequency in the reference corpus (Lee & Chen, 2009: pp. 152-153). Keywords are further described in Chapters 3 and 4.

software tools. Among their findings were that overused words can sometimes be attributed to the essay, i.e. are topic-related (Lee & Chen, 2009: pp. 153-154). Although they highlighted that topic-specific words may not necessarily be indicative of learners' problems in writing, overused function words such as *can* and *the*, however, are suspected to raise some concerns (ibid: pp. 153-154). As will be discussed in the next chapter, a keywords list acts as a good starting point for the examination of learner language in the present thesis, and it will be interesting to see whether much of learner writing is influenced by the text-type or by essay topic.

One neglected feature, however, is in the use of examining range (defined in this thesis as the distribution of items across different texts in the corpus, as explained in Chapter 3). Although nearly all LCR studies employ descriptive statistics via examining frequencies one way or the other, observing the distribution of items across a variety number of texts is less common, although crucial, as it indicates how widespread a particular linguistic phenomenon might be. Therefore, in the present thesis, classic methods such as examining frequencies of words in a corpus, is balanced with identifying range, which will be described further in the next chapter.

It is also important to combine quantitative and qualitative methods. As a typical example, De Cock (2011) compared the use of frequently recurring positive and negative evaluative adjectives between native speaker speech and in the spoken productions of advanced EFL learners from Chinese, French and German mother-tongue backgrounds. By looking at the distribution of these items alone, the study illustrates that there are more positive evaluative adjectives in both native and learner speech corpora overall (De Cock, 2011: pp. 202-203). However, De Cock cautions that judging merely on frequency counts can mislead researchers about the possibility that "both positive and negative evaluative adjectives can occur in non-assertive contexts" (e.g. *not too/that/so/as bad* and *not very/particularly good*) (2011: p. 203), hence it is suggested that a first analysis of the distribution of lexical items in the corpus is essential for further exploration on learner language. The present study, in turn, also makes extensive use of concordancing for qualitative analysis, to identify the way words are *used in context* (or *co-text*).

Of the various learner corpus studies that have identified particular linguistic categories, those that focus on verbs, modals and pronouns are most pertinent to the present thesis. Granger and Paquot (2009) explored the connection between learners' use of verb forms and verb lemmas in academic discourse with regard to ICLE. Following Biber et al. (1994), Granger and Paquot (2009) discovered that the under- and misuse of verbs such as *include*, *report* or *relate* by non-native learners may be contributed by the different registers of writing that were being compared. In their study, ICLE consists of learners' argumentative writing that is less representative of academic discourse features, such as citing sources and referring to tables and graphs. Another study (Salazar & Verdaguer, 2009) discovered that learners (Spanish in particular) over-use verbs such as *think*, *seem* and *know* in their argumentative writing when compared to texts written by American students. Learners also appeared to have difficulty in using polysemous verbs such as *feel* and their abstract meanings. Other corpus studies (e.g. (Altenberg & Granger, 2001; Granger & Rayson, 1998) focussed on high-frequency verbs such as *make* and *think*, which are less characteristic of academic writing.[17] Such studies show that the analysis of learner corpus data and their comparison with data from native corpora have highlighted some problems experienced by learners, e.g. lack of register awareness, phraseological infelicities, and semantic misuse.

Modals are also frequently investigated in the learner corpus literature (e.g. Gabrielatos & McEnery, 2005; Hyland & Milton, 1997; Römer, 2004). More specifically, Gabrielatos and McEnery (2005) discovered that learners use fewer epistemic modals such as *would* and *may* compared to their use in MA dissertations written by native speakers. This was also found in Hyland and Milton (1997), who conclude that the Chinese secondary school students in their study demonstrated a higher degree of assertiveness, or commitment to their statements, than the native-speaking students (Gabrielatos & McEnery, 2005: p. 325). On a different note, Römer (2004: p. 193) investigated the distribution of modals in a German English language textbook, and found that the occurrences of two modals (*can* and *will*) were significantly more frequent compared to modals *would*, *could*, *should*, and *might*. It was concluded that modals presented in English lessons and pedagogical materials in Germany "differ considerably from the use of those verbs in contemporary spoken British English" (Römer, 2004: p. 197). Although a specialised corpus was examined (textbooks), this corpus-driven study demonstrates how modal verbs are presented in language resources, which may consequently have implications for learner language production. [18] Similar research was found in Malaysian LCR, particularly work done by Mohamed Ismail et al. (2013); Mukundan et al. (2013); and

[17] Granger and Rayson (1998) noted that learners produce more infinitives rather than participle forms (i.e. past participles, -*ing* participles), which is more indicative of speech than academic writing.

[18] There are also Malaysian corpus studies that explore modal verbs usage specific to Malaysian English textbooks (e.g. Khojasteh & Kafipour, 2012; Mukundan & Khojasteh, 2011).

Vethamani et al. (2010). However, these local studies were not conducted using a contrastive approach; hence, this is an area in which the present thesis makes a contribution.

In addition to modality, pronoun use is also examined in learner writing (e.g. Breeze, 2007; Gilquin & Paquot, 2008). Breeze (2007: p. 19) found that Spanish learners over-use the personal pronoun *I* compared to in the native-speaking reference variety. Use of the personal pronoun *I* and its relative form *me* was also reported in Gilquin and Paquot (2008: p. 48), in which learners from various backgrounds (e.g. Chinese, Dutch, Finnish, Japanese) tended to employ more direct personal expressions such as *it seems to me* and *I would like*/*want*/*am going to talk about*… than in the native-speaking reference variety. However, Hyland (2002b: p. 354) discovered, in his study of identity in Hong Kong learner writing, that learners avoided using pronouns *I* and *we*. This was mainly because they believed that the pronouns "were inappropriate in academic writing, having been taught not to bring their own opinions into their texts" (ibid: p. 353). In regard to the Malaysian context, it will be interesting to find out whether Malaysian learners face the same situation, as to the best of the researcher's knowledge, there has not been any study done in this area. Essentially, to a great extent, description and evaluation of learner writing were based on the notion that learner essays feature high writer/reader visibility (McCrostie, 2008), and that learner writing exemplifies speech written down (Luzón, 2009).

One similar observation among the above-mentioned studies is that linguists have also employed functional analysis with regard to the examination of linguistic items. Studies such as Luzón (2009) and Römer (2004) show how investigating meanings of pronouns and modals in texts enhances the study, mainly in explaining descriptions about frequency counts and statistics. In addition to the use of computational statistics, the present study also employs functional analyses, by classifications of items according to their discourse functions or meanings (further described in Chapter 3). This, in turn, adds to the qualitative approach promoted in the thesis, in which linguistic items are discussed in relation to their use in context. Contrary to the previous section, which discussed how learner corpus studies have looked at word-based and category-based analyses, the next section will present a discussion on ways to explore phraseology, by reference to existing scholarship.

2.4.2 Multi-word combinations and Phraseology

The study of phraseology has grown in recent years, along with recognition of its importance in both applied linguistics and learner language (Byrd & Coxhead, 2010; Paquot & Granger, 2012). Phraseology, as defined by Howarth (1998: p. 24), is "the study of word combinations",[19] including the use of collocations. Granger and Paquot (2008: p. 28) highlight that there are two major approaches to phraseology: the traditional approach (following Cowie's 1998 phraseological continuum); and the frequency-based (or distributional) approach, which they state is the result of an inductive approach, influenced by Sinclair (1987). In the 'traditional' approach, word combinations are distinguished along a continuum "which goes from free combinations to pure idioms through restricted collocations and figurative idioms", mainly via a top-down classification on the basis of linguistic criteria (Granger & Paquot, 2008: p. 28). In the more recent, frequency-based approach to phraseology, a bottom-up corpus-driven approach is used to identify lexical co-occurrences (i.e. word combinations). This approach tends to use automatic quantitative analyses to identify features such as n-grams/lexical bundles or collocations, including analysis of strength of association between pairs of words (collocation) and frequent phrases (Stubbs 2003, Peromingo 2012). In this thesis, the frequency-based approach to phraseology is followed, allowing for the extraction of recurrent continuous sequences of two or more words, viz. "recurrent expressions, regardless of their idiomaticity, and regardless of their structural status" (Biber et al., 1999: p. 990).

Much LCR that can be classified as phraseological focuses on the analysis of collocation, or the co-occurrence of words. While collocation can be defined in different ways (see McEnery & Hardie, 2012 for an overview), many corpus linguists adopt a statistical approach to collocation and refer to it as the non-random co-occurrence of words (Xiao & McEnery, 2006: p. 105). Biber et al. (1998: p. 84) define collocation as "words [that] have strong association patterns with other words", while Stubbs (2003: pp. 226-227) defines it as "the habitual co-occurrence of two unordered content words, or of a content word and a lexical set". The investigation of collocates in the present thesis is in line with the approach labelled 'collocation-via-significance'.[20] This means that collocates are determined by applying statistical tests (e.g. MI-score, T-score, etc.), which "compare the frequency of each word

[19] However, one must remain aware of "the highly variable and wide-ranging scope of the field and the vast and confusing terminology associated with it" (Granger & Paquot, 2008: p. 27).

[20] McEnery and Hardie (2012: pp. 126-130) make note of the two general definitions of investigating collocations: 1) collocation-via-concordance, and 2) collocation-via-significance.

within the window of text defined by the span around the node word,[21] against its frequency in the rest of the corpus" (McEnery & Hardie, 2012: p. 127). In doing so, collocates are not simply determined on the basis of co-occurrence within a given span, it is also determined that their association is non-random, i.e. not due to chance. As will be shown in Chapter 3 (Section 3.4.3), both function words and lexical words are explored.

Collocation is one of the most important discoveries in CL studies, as "[t]here are countless combinations of words which are grammatically possible but do not occur, or occur only rarely, [and] there are collocations with very similar meanings which occur with great frequency" (Cook, 2003: p. 73). Collocations are also regarded as an important part of native speaker competence (Nesselhauf, 2003: p. 223). According to Sinclair (1991: p. 115), collocation illustrates the idiom principle.[22] However, Granger (1998) hypothesizes that learners make use of the open-choice principle (see Sinclair, 1991: p. 109-110), [23] as opposed to native speakers who tend to operate more according to the former principle. This suggests that learners have a tendency to learn English words in isolation instead of in chunks, which in turn becomes detrimental to learners, as English has many of these combinations, and some can be more arbitrary than others.[24] Collocations, therefore, are viewed by scholars as one of the most complex aspects in corpus linguistics, and a necessary component of L2 lexical competence, in addition to achieving native speaker fluency (Sinclair, 1991; Cook, 2003; and Nesselhauf, 2003).[25]

Granger (1998) investigates the two concepts, collocations and formulae, or pragmatic idioms that could prove to be a constraint for learners on achieving near-nativeness in the target language. She concludes that, in order to acquire knowledge of a given lexical phrase, both chunks of words and their contextual functions need to be learned (Granger, 1998: p. 157). As

[21] A node is the target item under investigation whereas collocates are items that appear surrounding it within a specified span/window (Sinclair et al., 2004: p. 10).

[22] The idiom principle accounts for a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments. For example, *of course* is usually understood as a phrase (i.e. single word) rather than perceiving *of* as a separate preposition, according to Sinclair (1991: p. 115).

[23] The open-choice principle describes how sentences are created on a kind of slot-and-filler basis: "the grammar generates the slots and then any item which fits syntactically and semantically into a given slot may, in principle, be used" (Butler, 2004: p. 155).

[24] Many studies have also found that deviant collocations made by learners are often a cause of interlingual transfer (e.g. Fan, 2010; Laufer & Waldman, 2011; Nesselhauf, 2004).

[25] As mentioned by Howarth (1998: p. 31), learners face problems in achieving near-nativeness mainly from inappropriate selection of conventional phraseology, and that learners face difficulty between free and restricted combinations of words. In addition, since learning in chunks does not always apply, Howarth asserts that "it is far more efficient to teach the nature of the phenomenon and [develop awareness]" (1998: p. 42) of word combinations to learners at a much earlier stage.

Nesselhauf (2003: p. 226) notes, word combinations can be distinguished into three categories, free combinations, collocations, and idioms. However, she points out that the delimitations of these classes are almost impossible, as “word combinations differ along a scale” (ibid: p. 226), and that the concept of collocations should be treated with the notion of ‘restricted sense’. While this may be problematic, Walker (2011) suggests that learners can produce collocations better by firstly understanding how word combinations were formed, instead of trying to memorise large numbers of collocations and being highly dependent on dictionaries. He argues that it is “a process which can be partially explained by examining some of the linguistic features and processes which influence the way collocations are formed” (ibid: p. 292).

In response to Walker (2011), the present study makes use of collocation analysis, particularly in highlighting and comparing the ways in which different words are selected/used by different writers. Differences between learners’ collocation and the reference language variety thus, illustrate how word meanings are derived by their collocational patterns (Hunston, 2002b).

As Hasselgren (1994) argues, corpus analyses are better suited to examining more of learners’ types of wrong word choice that lead to wrongness, not of meaning but rather of collocation or style. She states that “learners cling to the familiar L1 vocabulary boundaries and impose them on the L2 [i.e., second language], which results in a false one-to-one translation equivalence”, also known as “lexical teddy bears” (Hasselgren, 1994: p. 256). For example, she notes that “the wide collocational range of the Norwegian L1 word *styre* has been mistakenly assigned to the L2 cognate[26] *steers* in the clause *the time schedule no longer steers their activities*” (Hasselgren, 1994: p. 243). In another study, Altenberg and Granger (2001) found that Swedish and French learners tend to make collocational mistakes involving the grammatical and lexical patterning of *make*, specifically the underuse of de-lexical uses (e.g. *make a balance* instead of *strike/find a balance*) and causative uses of the verb (e.g. N *makes the air polluted* instead of N *pollutes the air*). In some instances, misuses of the phrasal verb *make* are attributed to positive transfer from learners’ L1 constructions. De Cock and Granger’s (2004: p. 241) findings demonstrate this in their investigation of the high-frequency verb *make* in learners’ dictionaries, and they discovered that learners need to be aware of “the highly

[26] A cognate is defined by Crystal (1991: 60) as “a linguistic form which is historically derived from the same source as another form”. According to Hasselgren (1994), cognates may be perceived as equivalent to the learner, and therefore potentially problematic. Cognates are regarded as ‘false friends’ and “lie behind such errors as the use of crib (from Norwegian krybbe) for cradle” (Hasselgren, 1994: p. 240).

polysemous and phraseological nature of high frequency words" that learners tend to easily misinterpret if translated into their L1 equivalent. Similarly, past research has shown that Malaysian learners face problems with phrasal verbs, particularly those involving the verbs *pull*, *come*, *go*, *get*, and *look* (Akbari, 2009; Zarifi & Mukundan, 2012). In summary, LCR on collocation has shown that both EFL and ESL learners often struggle with producing word combinations adequately.

This leads to the examination of lexical bundles or n-grams, using corpus methods. 'N-grams' or other terms like 'clusters' and 'bundles' are often used more or less interchangeably in the literature in reference to multi-word sequences or recurrent word combinations (Byrd & Coxhead, 2010). I shall be using the term *lexical bundle* in this thesis. Lexical bundles, which are defined by Biber and Barbieri (2007: p. 264) as "most frequently recurrent sequences of words" [their frequency threshold is at least 40 times per million words], are "words which follow each other more frequently than expected by chance, helping to shape text meanings and contributing to our sense of distinctiveness in a register" (Hyland, 2008: p. 5). While there are various definitions of such sequences, in this thesis lexical bundles are conceptualised similar to the way 'n-grams' are theorised by Stubbs (2003). A lexical bundle is hence defined as "a recurrent 'chain' of word-forms. A 'chain' is defined here as a linear sequence of uninterrupted word-forms, either two adjacent words, or longer strings, which occur more than once in a text or corpus" (Stubbs 2003: p. 230). As explained in Chapters 3 and 6, the investigation of lexical bundles in the present study involves recurrent chains of three to four words, with a minimum frequency of five (Section 3.4.2). While the investigated lexical bundles all occur more than once in the node corpus (MCSAW), some occur only once in the reference corpus (LOCNESS).

By identifying and comparing lexical bundles using corpus tools, research has shown that it becomes possible to examine non-idiomatic phrases that are indicative of learner language (or learner writing style). Furthermore, bundles are often purposeful for textual cohesion, and thus, analysis of lexical bundles, in turn, can provide explanations for non-nativeness in learner writing. LCR studies that focus on the investigation of lexical bundles are many. These include studies by Ädel and Erman (2012), Chen and Baker (2010), Ebeling (2011), and Peromingo (2012). In general, linguists have found that learner writing exhibits less variety of lexical bundles in the learners' repertoire when compared to native-speaking or expert English writers (Ädel & Erman, 2012; Chen & Baker, 2010). According to Ädel and Erman (2012: p. 86), learners are considered “less mature academic writers” because they present “greater use of anticipatory ‘it’ constructions, coupled with relatively informal lexical choices, involving ‘hard’ and ‘easy’ for these constructions” (ibid: p. 86). In addition, Peromingo (2012) found that bundles in learner writing have also been linked to a potential influence of learners’ mother tongue. More importantly, findings have been “shown to be largely similar to those of the phraseological research tradition in SLA” (Ädel & Erman, 2012: p. 81)[27] and therefore, is another good reason to explore lexical bundles using corpus methods. In addition, Ebeling (2011) discovered that the identification of bundles can represent particular text-types. By examining lexical bundles and their discourse functions, Ebeling argues that “[t]he method explored […] paves the way for similar studies on text-types across more and different kinds of corpora” (ibid: p. 69).[28] Following this, the present research includes the investigation of lexical bundles and their discourse functions, aside from merely examining single-word lexical items (as mentioned in Section 2.4.1).

While categories of word combinations can be tricky to identify, and classify (Granger and Paquot 2008), it is still a very important fact of learner language that is better understood with the incorporation of corpora (Cheng et al., 2008). In keeping with the contrastive nature of the research, in this thesis, investigation of lexical bundles is firstly compared across their use in two corpora, which reveals differences or similarities in the patterns writers make in their essays. Salient bundles in the learner corpus are extracted using WordSmith Tools, which are further described in Chapter 3. Bundles are then explained in more detail via functional analyses based on their discourse functions, which, as discussed in Chapters 3 and 6, are adapted following taxonomies offered in the literature (e.g. Biber et al., 2004; Chen & Baker, 2014).

In the present research, both quantitative and qualitative methods are considered, in that the investigation of linguistic phenomena in Malaysian learner English writing are examined both in terms of individual lexical items as well as through the examination of lexical bundles. However, as will be discussed in Chapter 3, this research does not include other corpus

[27] Some of the findings in SLA research show that learners under-use collocations compared to native speakers in writing, over-use high-frequency collocations, have poorer intuitions about (a)typical collocations, and take 30% longer to make judgements regarding collocational frequencies (Ädel & Erman, 2012: p. 82).

[28] One example is a comparative analysis of recurrent word-combinations (or lexical bundles) in linguistics and business texts produced by Norwegian learners of English and native speakers of English, in Ebeling and Hasselgård (2015).

techniques such as tagging, parsing or annotating, because it follows the corpus-driven approach (Tognini-Bonelli, 2001).

2.4.3 Connection between novice writers and evidence of spoken English forms in writing

In addition to LCR studies that explore various linguistic phenomena, a number of studies have examined evidence of spoken English forms in novice writing, particularly in the writing by non-native English speakers (McCrostie, 2008; Gilquin & Paquot, 2007; Cobb, 2003). In this section, a brief review is presented of this work, as it will be drawn upon in interpreting findings in this thesis.

The scholarship in LCR has shown that learners tend to use spoken features in their writing (McCrostie, 2008; Gilquin & Paquot, 2007). The impression is that learner writing (even that of advanced learners) resembles speech written down. For example, “[h]eavy use of personal pronouns in an academic or professional text creates a style that is perceived as too simple, or too similar to speech” (Breeze, 2007: p. 15). Cobb (2003: p. 395) argues that learners rely more on “the restricted, context-determined lexicon of spoken language rather than deploying the broader lexicon typical of [native-speaking] writing”. Table 2.1 presents some of the spoken-like items found to be over-used in argumentative essays written by learners from the ICLE corpus (Gilquin & Paquot, 2008: p. 51).

Table 2.1: Spoken-like overused lexical items per rhetorical function

Rhetorical function	**Spoken-like overused lexical item**
Exemplification	*like*
Cause and effect	*thanks to* *so* *because* *that/this is why*
Comparison and contrast	*look like* *like*
Concession	sentence-final adverb *though*
Adding information	sentence-initial *and* adverb *besides*
Expressing personal opinion	*I think* *to my mind* *from my point of view* *it seems to me*

Expressing possibility and certainty	*really* *of course* *absolutely* *maybe*
Introducing topics and ideas	*I would like to/want/am going to talk about thing* *by the way*
Listing items	*first of all*

One explanation for this phenomenon is the tendency for learners to use more interpersonal involvement (Cobb 2003: 395), i.e. expressing the writer's attitude to a message (typically realised through use of modals but see also the category 'expressing personal opinion' in Table 2.1). As McCrostie notes, "[i]nterpersonal involvement carries the signalling load in spoken language, whereas message content carries the load in written language" (2008: p. 99). In a study of Malaysian learner writing, Vethamani et al. (2010) found that evidence for features of colloquial Malaysian English in writing was related to learners' use of compensation strategies and simplification features, in overcoming their limitations in the target language. Hence, it will be interesting to examine in this thesis if traces of spoken English forms are also present in the Malaysian argumentative written corpus, MCSAW.

2.5 Summary

To summarise, three things have been described with regard to the aim of the review. The chapter firstly highlighted how the scholarship has investigated learner language, in which there are two kinds: non-contrastive and contrastive studies. Following the latter kind, the chapter then described Granger's (1998) Contrastive Interlanguage Analysis (CIA) approach, which will underpin the present thesis, i.e. comparing learners' writing in the Malaysian corpus, MCSAW, against a comparable reference language variety, LOCNESS. As the review has also shown, contrasting two sets of argumentative writings (MCSAW vs LOCNESS) not only enables us to uncover empirical evidence into descriptions of learners' argumentative writing relative to a comparable reference variety, but also fills a gap, methodologically, in terms of the dearth of Malaysian contrastive LCR.

The final part of the chapter then demonstrated how past CIA research has explored various linguistic phenomena, namely lexical, grammatical and collocational features, which are important in analysing learner language. A brief review of work done on the connection between novice writers and evidence of spoken English forms in writing was also presented. In summary, this thesis sets out to extend from past LCR studies, and examine Malaysian learner language via an investigation that combines both individual lexical items and recurrent lexical bundles. This is important, so as to provide a wider spectrum of findings that also account for the relationship between lexis and phraseology in learner writing. In addition to the use of computational statistics, the present study makes use of functional analyses by classifications of items according to their use in discourse, and in turn, responds to the need for more qualitative results. Details of the methodology are further explained in the next chapter.

Chapter 3: Data and Methodology

3.1 Introduction

The aim of this chapter is to outline the methodological framework used in this thesis in exploring learner language as well as describing the data (i.e. corpora) that will be the basis of this study. The chapter contains four main sections: Section 3.2 of this chapter will briefly discuss key debates in the field of CL and the relevant decisions taken; while Section 3.3 provides a comprehensive description of the corpora (Sections 3.3.1 and 3.3.2), the argumentative essay (Section 3.3.3), the software (Section 3.3.4), followed by the step-by-step procedure (Section 3.4), from the selection of items to analysis of keywords, collocates and concordances.

By clearly describing the methods of practice in this thesis, it is hoped that transparency of the research can be maintained, for others to be able to evaluate the work and replicate if necessary. This is important because "visibility is the necessary condition for replicability, as far as replicability is achievable" (Marchi, 2013: p. 71). Moreover, our achieving a sense of confidence in the validity of CL as an "empirical, scientific enterprise" (McEnery & Hardie, 2012: p. 17) relies heavily in the total accountability of our data and the process of checking and rechecking them to "meet the standard of falsifiability [and] replication" (ibid: p. 17).

3.2 Key debates and decisions

One debate concerning CL is about its status: is it a methodology or a theory? Some linguists find the corpus not merely to be a tool of linguistic analysis but an important area in linguistic theory (Stubbs, 1993; Teubert, 2005). Others have argued for CL to be a major *methodological* paradigm in applied and theoretical linguistics (Gries, 2006; Tognini-Bonelli, 2001; Mahlberg, 2005). Thompson and Hunston (2006) go on to add that CL is a methodology that can be aligned to any theoretical approach to language. One point worthy to make is that "[c]orpus linguistics may be viewed as a methodology, but the methodological practices adopted by corpus linguists are not uniform", and are driven by theoretical considerations (Taylor, 2008: p. 181). In other words, CL can be considered "a methodology innovation" (Lee, 2008: p. 87), which comprises "a set of theoretical positions and beliefs about the nature of language and how we can study it" (ibid: p. 87). In sum, CL provides a plethora of approaches to the study of language (McEnery & Hardie, 2012). In this thesis, CL is treated as a method for the analysis of language, in which the corpus acts as a machine-readable written language sample that has been assembled in a principled way to be further analysed and discussed for the purpose of linguistic research.

Another crucial debate concerns the distinction between corpus-based and corpus-driven linguistics,[29] which relates to the decision of how one approaches the data, or "the degree to which empirical data from a corpus is relied on" (McEnery & Hardie, 2012: p. 151). While it is beyond the scope of this thesis to provide an overview of the different ways in which corpus-based and corpus-driven linguistics have been conceptualised, a key contribution is that by Tognini-Bonelli (2001). She states that, within the corpus-based approach, "the corpus is used to validate or exemplify existing theories" (Tognini-Bonelli, 2000: p. 236), and therefore theoretical statements would be considered the starting point of linguistic analysis. The corpus-driven approach, on the other hand, starts from the evidence (i.e. corpus data) and tries to account for it (ibid: p. 236). As a result, findings are inferred from the analyses of data (Butler, 2004). In other words, the corpus-driven approach constitutes a bottom-up approach to linguistic analysis, where corpus data are the starting point:

> In a corpus-driven approach the procedure to describe the data is therefore inductive in that it is statements of a theoretical nature about the language or the culture which are arrived at from observations of the actual instances. The observation of language facts will lead to the formulation of a hypothesis to account for these facts; this in turn will lead to a generalisation based on the evidence of the repeated patterns in the concordance; the last step will be the unification of these observations in a theoretical statement (Tognini-Bonelli, 2000: p. 207).

Corpus-driven research also tends to work with 'raw' text data, rather than annotated corpora, although this is not always the case (see Granger & Rayson 1998 for a corpus-driven study of part-of-speech tagged data). While corpus-based research implies that one approaches the data from a preconceived notion or theory (or with a hypothesis in mind), in the corpus-driven approach, the data is examined without any preconceptions at all (Tognini-Bonelli, 2001: p. 84). In particular, the corpus-driven approach to linguistic investigation has demonstrated that

[29] Other terminologies include deductive and inductive approaches, and top-down and bottom-up approaches. Matters get more complicated because the term *corpus-based* is sometimes simply used as a cover term for any linguistic study that is based on analysis of a corpus regardless of the approach to the data (i.e. any kind of corpus-informed or corpus-inspired research) (McEnery & Hardie, 2012: p. 6).

much of the language we use is made up of semi-prefabricated chunks of language (Biber, 2009: pp. 300-301). As Butler (2004: p. 175) asserts, advocating inductive procedures, i.e. the corpus-driven approach, “has certainly revealed a great deal of important information about the way in which naturally occurring language is organised”. In reality, it appears that most research is neither strictly corpus-based nor strictly corpus-driven and there are clear differences in how these terms are used and applied in contemporary corpus linguistics. In this thesis, the term ‘corpus-driven approach’ is used very broadly to refer to inductive or bottom-up research in which corpus data is used as the starting point in investigating patterns of overuse in language writing.

Many of the learner corpus studies described in Chapter 2 apply a corpus-driven approach (in a very broad sense, i.e. an inductive approach). This is also the approach taken in the present thesis, which I would position as more corpus-driven than corpus-based, since it does not start with a specific hypothesis that is tested, and uses a raw text corpus rather than a tagged corpus. At the same time, some of the analyses in the present research are informed by frameworks developed in past research, and so are not entirely theory-free. As explained in more detail later, I start with analysis of keywords (including classification of types of keywords, examining keyness, range, and key keywords), followed by a comparative analysis of the selected keyword (investigating relative frequencies between corpora, frequency across L1 backgrounds, and dispersion plot). Then, I continue to investigate collocates of the keywords, i.e. words co-occurring with the keywords, using two statistical measures (t-score and MI-score); then ending with the qualitative analysis via use of concordancing. Linguistic frameworks that are drawn on in the analysis include classification schemes for the analysis of n-grams (Biber et al., 2004; Chen & Baker, 2010) and discourse functions (Biber et al., 2004; Chen & Baker, 2010; Bednarek, 2008b).

In sum, Adolphs and Lin (2011: p. 597) state, CL is essentially concerned with language use in real contexts, often through empirically based linguistics and data-driven description of language. In other words, “it is an empirical approach to the description of language use; it operates within the framework of a contextual and functional theory of meaning; [and] it makes use of the new technologies” (Tognini-Bonelli, 2000: p. 206). This gives a benefit to CL studies, in that they are “intrinsically more verifiable than introspectively based judgements” (McEnery & Wilson, 2001: p. 14). As discussed in Chapter 2, patterns of learner language are best explored via contrastive corpus studies. In this thesis, Malaysian learner argumentative writing will be contrasted with a reference language variety. The next section, thus, introduces the two relevant corpora as well as the software tool used to analyse them.

3.3 Data and software

Corpora, as Adolphs and Lin (2011: p. 597) note, "are designed to represent a particular language variety". In the present study, two English varieties are compared and contrasted, using a set of tools provided by a computer software. The following sub-sections offer a brief overview of the corpora used in the present study, and the tools used to conduct the contrastive, corpus-driven study.

3.3.1 Malaysian Corpus of Students' Argumentative Writing (MCSAW)

As mentioned in Chapter 2, learner corpora contain collections of texts produced by learners of a language. More specifically, a learner corpus can be used as "a basis for better descriptions of different varieties that emerge from communication between speakers who communicate in a language other than their first language" (Adolphs & Lin, 2011: p. 599). Thus, the investigation of learner corpora allows linguists to identify patterns in a particular variety of learner English, and to compare the language of the learner to that of other users of a language, i.e. via Contrastive Interlanguage Analysis (CIA). The learner data comes from the second version of *Malaysian Corpus of Students' Argumentative Writing* (henceforth, MCSAW) (Mukundan & Kalajahi, 2013), which contains 565,500 words of argumentative essay writing by 1460 students from schools and colleges in four states of Malaysia (Selangor, Negeri Sembilan, Melaka and Kelantan). These data are made up of students from three different levels: Form Four (16-year olds), Form Five (17-year olds), and College students. However, only first-year students were involved in the collection of data for the College files.

The data in MCSAW are untagged, but include information on the writers' first language, i.e. their L1 (Mukundan & Kalajahi, 2013). These are provided in the manual as 'metadata', or data about data, such as descriptions of technical specifications and data, data collection procedures, and demography of students. It is also noteworthy that the Malaysian writers comprise learners with three main L1 backgrounds, namely Malay, Chinese and Indian. While the MCSAW compilers did not specify the learners' level of proficiency (Mukundan &

Kalajahi, 2013: p. 4), it can be implied that the three educational levels, namely Form 4, Form 5 and College level, should have sufficient grasp of the language since these are upper secondary school and college students that have gone through at least 9 years of ELT experience in school. In addition, although the three different learner groups possess different types of L1, it is not possible to generalise that a particular ethnic group uses a particular feature because it is associated with their L1; rather this may be an issue that is related to competency.

The students were assigned two essays, entitled 'Do you think Facebook has more advantages than disadvantages? Discuss your reasons', and 'What are the advantages and disadvantages of living in a hostel'. They were then asked to write a 250-word argumentative essay on one of the topics during class time. The compilers' reasons behind assigning the specific topics were not only for their familiarity to the students but also their capability of stimulating the students to write more productively (Mukundan & Kalajahi, 2013: pp. 5-6). According to the compilers, MCSAW was formed to serve as "a baseline data of the Malaysian students' English language proficiency in written forms and also to study developmental patterns through the data gained" (Mukundan & Kalajahi, 2013: p. 2). However, the present thesis focusses only on one level (college students' writing), and therefore developmental patterns across proficiency levels are not studied in the thesis. Nevertheless, MCSAW was chosen mainly because it is more specialised: texts in MCSAW belong to a particular type, i.e. argumentative essays. Other Malaysian corpora either include Malaysian English rather than learner language (e.g. the Malaysian sub-corpus of International Corpus of English, ICE), or consist of only school children essays (e.g. the English of Malaysian School Students, EMAS). This, in turn, motivated the use of MCSAW in the thesis, consisting of a relatively current collection of argumentative essays by Malaysian advanced-level students to be compared against the reference corpus.

3.3.2 Louvain Corpus of Native English Essays (LOCNESS)

As mentioned earlier, contrastive corpus studies in LCR compare and contrast two or more corpora, usually with one being the reference language variety. The *Louvain Corpus of Native English Essays* (henceforth, LOCNESS) is used as the reference corpus in this thesis, because "it contains argumentative essays produced by [novice] English-speaking writers and is therefore arguably a more reliable basis for comparisons with learner corpora like [MCSAW] than more general corpora (e.g. the British National Corpus, BNC)" (Granger 2015: p. 17).

The reference corpus used is a compilation of texts from LOCNESS, compiled at the Universite Catholique de Louvain (UCL) by Sylviane Granger and her team. Currently, the total number of words is 324,304.[30] LOCNESS contains a mix of argumentative and literary essays from various topics, written by British A-level students and British and American university students. As a process of delineation, only texts from the A-levels, BRSUR, and USARG sub-corpus are used, because they contain the most argumentative essays.[31]

Following Scott and Tribble (2006; p. 133), both corpora are described as novice writing samples, since they contain "unpublished pieces of writing that have been written in educational or training settings". Consequently, LOCNESS does not act as the native speaker norm (Granger, 2015: p. 17), since speakers of LOCNESS do not represent the larger group of English first-language speakers, and since they are novice writers. The difference between using something as a reference language variety and using something as 'the native speaker norm' is in response to criticisms about learners having native speaker norms as a target (Hunston, 2002a: p. 211). By using LOCNESS as a reference language variety rather than the 'norm', such criticisms can be addressed. For the purpose of the present study, LOCNESS is compatible for a CIA to be conducted with MCSAW mainly due to two reasons: comparable writer authorship, and genre of writing.

As pointed out in Section 3.3.1, only Malaysian college learners from MCSAW will be analysed. This is in order to reach near-comparability to the standard age group of essays collected in LOCNESS (i.e. advanced-level students). In addition, all argumentative-type essays from LOCNESS are included, regardless of topic, so as not to be too specific, as well as allowing for more running words in the reference corpus. As noted earlier, essays from MCSAW-College files are only written by first-year students, hence findings of the analysis would not be generalised to all Malaysian college students. Table 3.1 below illustrates the two corpora that will be used in the thesis:

[30] As of 24th September 2013, LOCNESS can be obtained and requested from the website Learner Corpus Association http://www.learnercorpusassociation.org/resources/corpora/locness-corpus/.

[31] In LOCNESS, one text file does not correspond to one essay. Rather, one text file contains several essays, in some cases including both literary and argumentative texts.

Table 3.1: MCSAW and LOCNESS

Corpus	Demographics	Topics	Word count
MCSAW (target corpus)	Malaysian second language learners from college educational level	Consist of only two: • Do you think Facebook has more advantages than disadvantages? Discuss your reasons. • What are the advantages and disadvantages of living in a hostel?	197,293 running words in overall corpus
LOCNESS (reference corpus)	Argumentative essays written by British A-level and American and British University students	Argumentative essays range from a number of topics, such as: • 'A single Europe: A loss of sovereignty for Britain' • 'Great inventions and discoveries of 20th century and their impact on people's lives' • 'Money is the root of all evil' • Abortion • Capital punishment • Euthanasia • Gender roles in our society • Legalisation of marijuana • Parliamentary system • Recycling • Transport	323,929 running words in overall corpus

By comparing MCSAW with LOCNESS, it is possible to employ the CIA approach, that is, in investigating the Malaysian learner English variety against a native-speaking English variety. In addition, the contrastive approach is deemed comparable in terms of argumentative writing genre, as well as for the reason that writers of both corpora are similarly representative of being novice writers.

3.3.3 The argumentative essay

Since both corpora contain argumentative essays, it is important to briefly describe this genre. In the area of genre studies, argumentative essays have long been studied, beginning with Veel (1997), followed by Lock and Lockhart (1998), and Derewianka (1990). However, according to Qian (2010: pp. 57-58), Hyland's (1990) argumentative essay analysis framework is most suitable in analysing essays written by non-native English speakers. Hyland (1990: p. 68) states that "the argumentative essay is defined by its purpose which is to persuade the reader of the correctness of a central statement", and is usually made up of three major parts: Thesis, Argument, and Conclusion.

According to Hyland, in the thesis stage, the proposition to be argued is firstly introduced. This process can be carried out through four ways: by including a controversial statement as an 'attention grabber'; by presenting background information on the topic; by providing a brief support of the proposition; or by introducing/identifying a list (Hyland, 1990: p. 69).

The argument stage, on the other hand, includes four types of argument sequences that "can be repeated indefinitely" (Hyland, 1990: p. 69). They are: signalling the introduction of a claim and relating it to the text; rephrasing/repeating the proposition; stating reason for acceptance of the proposition (either through the strength of perceived shared assumptions, a generalisation based on data or evidence, or by force of conviction); and supporting the proposition via explicating assumptions used to make the claim, or providing data (ibid: p. 69).

Finally, the conclusion stage, which synthesises the discussion and affirms the validity of the thesis, also involves four steps. These are: the 'marker', which signals the conclusion; the 'consolidation', which presents the significance of the argument to the proposition; the 'affirmation', which restates the proposition; and the 'close', which widens the context or perspective of the proposition (Hyland, 1990: p. 69). It is important to be aware of the structural conventions of this genre, since they may impact on learners' writing. While a full genre analysis is beyond the scope of this thesis, I will consider *where* in a text a particular linguistic feature tends to occur, using the plot function, described in the following section, which introduces the software tool used in this thesis (WordSmith Tools).

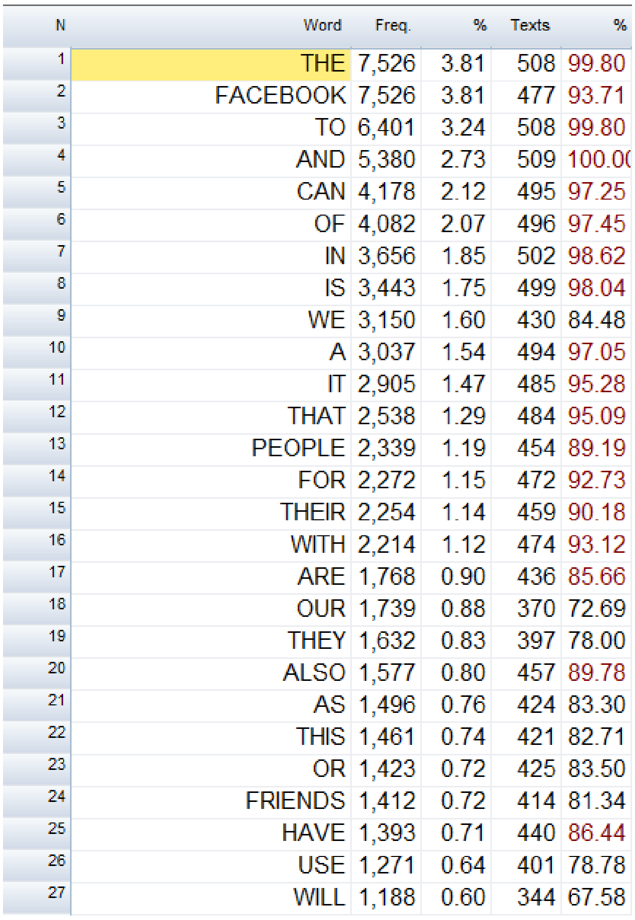
3.3.4 WordSmith Tools

A number of software packages are available that facilitate the manipulation and analysis of corpus data (e.g. AntConc, MonoConc). In the present thesis, WordSmith Tools is chosen. Since 1996, Mike Scott has developed a suite of tools that are published by Oxford University Press, named WordSmith Tools, for lexical analysis (Scott, 2008).[32] It includes a wide range of programs and functions, such as producing wordlists, keywords and clusters, as well as

[32] Version 6.0 was used for this analysis (Scott, 2012).

plotting distribution and showing collocational patterns. All word counts, keywords, collocates and general statistics provided in the present thesis are calculated using WordSmith Tools. The WordList tool, as the name suggests, creates word lists, arranging them by frequency and/or alphabetically. Table 3.2 is a sample of a word list taken from MCSAW.

Table 3.2: Sample of MCSAW word list



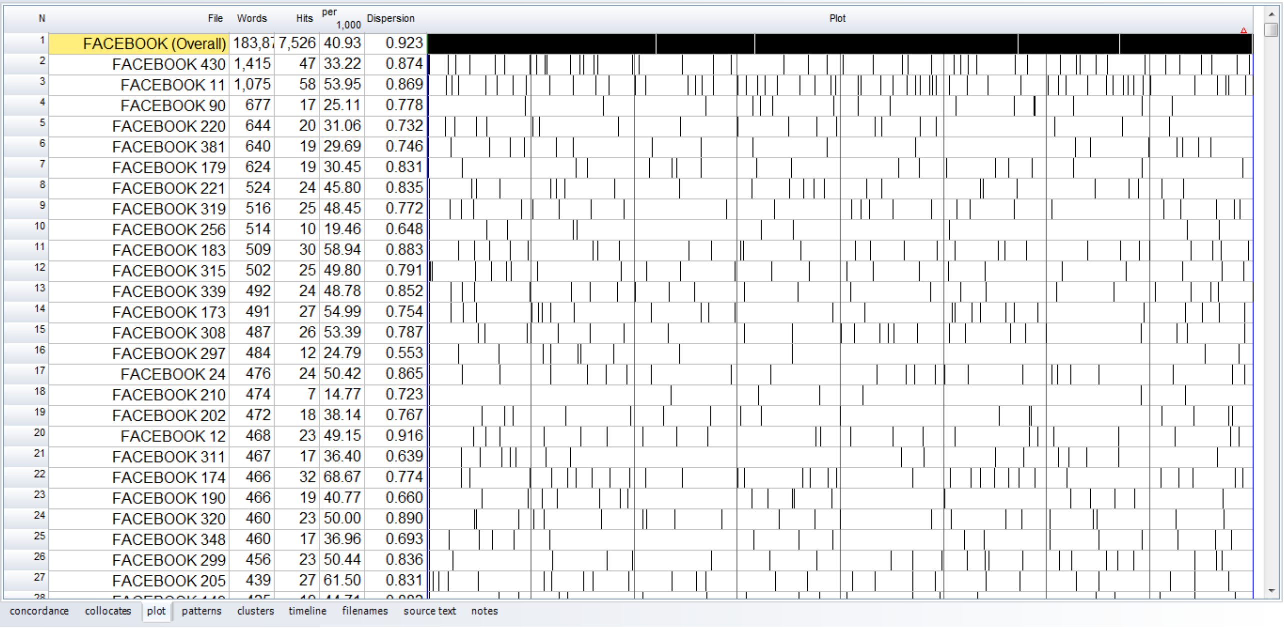
N	Word	Freq.	%	Texts	%
1	THE	7,526	3.81	508	99.80
2	FACEBOOK	7,526	3.81	477	93.71
3	TO	6,401	3.24	508	99.80
4	AND	5,380	2.73	509	100.0
5	CAN	4,178	2.12	495	97.25
6	OF	4,082	2.07	496	97.45
7	IN	3,656	1.85	502	98.62
8	IS	3,443	1.75	499	98.04
9	WE	3,150	1.60	430	84.48
10	A	3,037	1.54	494	97.05
11	IT	2,905	1.47	485	95.28
12	THAT	2,538	1.29	484	95.09
13	PEOPLE	2,339	1.19	454	89.19
14	FOR	2,272	1.15	472	92.73
15	THEIR	2,254	1.14	459	90.18
16	WITH	2,214	1.12	474	93.12
17	ARE	1,768	0.90	436	85.66
18	OUR	1,739	0.88	370	72.69
19	THEY	1,632	0.83	397	78.00
20	ALSO	1,577	0.80	457	89.78
21	AS	1,496	0.76	424	83.30
22	THIS	1,461	0.74	421	82.71
23	OR	1,423	0.72	425	83.50
24	FRIENDS	1,412	0.72	414	81.34
25	HAVE	1,393	0.71	440	86.44
26	USE	1,271	0.64	401	78.78
27	WILL	1,188	0.60	344	67.58

Word frequency information is useful to identify highly frequent words, which may indicate what a text is about (Scott, 2001). WordSmith Tools also lists the word frequencies according to the number of texts. This is called ‘range’, in which a word is calculated through its occurrence across a number of texts (under the column ‘Texts’). For instance, the list of 27 words in Table 3.2 shows that the most frequent word is *the*, which occurs 7,526 times, in 508 texts out of the total 509 texts. Notice that *Facebook* has the same frequency,[33] but only occurs in 477 texts, so it has a different range. This means that not only is it necessary to observe the frequency of a word, but the range also tells us the consistency with which it occurs across a number of texts. In turn, this signals the number of learners that use this word in their writing. More specifically, investigating range allows us to consider which linguistic features are commonly used by how many learners. In this thesis, range is an important feature for the interpretation of MCSAW findings; and one of the aims of this thesis is to test whether range is useful as a down-sampling technique in the context of LCR.

In addition, WordSmith Tools allows us to produce lists of clusters or bundles. As Stubbs (2003: p. 230) defines them, bundles are "a recurrent 'chain' of word-forms [i.e.], a linear sequence of uninterrupted word-forms, either two adjacent words, or longer strings, which occur more than once in a text or corpus". In this thesis, bundles are the focus of analysis in Chapter 6. Bundles are computed automatically via WordSmith Tools, which determines the number of words in a bundle (3 to 4-words) and their minimum of occurrence (5 times in MCSAW). As a reminder (see Section 2.4.2), lexical bundles are operationalised somewhat differently than in studies such as Biber and Barbieri (2007), and the operationalisation in this thesis is more akin to the definition of n-grams provided by Stubbs (2003). As will be described in Section 3.4.2, occurrences of bundles in MCSAW are then examined relative to their occurrences in the reference corpus, for comparison purposes.

Another feature available in the WordSmith suite of tools involves measuring the 'dispersion' of words. This shows where and how evenly the word is distributed within the text (at the beginning, middle or end), usually through viewing the 'dispersion plot', as shown in Figure 3.1.

[33] The fact that *Facebook* occurs with the same raw frequency as *the* is actually a rare coincidence given that the article *the* is considered the most frequent word in English (McEnery & Hardie, 2012). One possible reason for this is the underuse of articles in Malaysian English writing (Mia Emily et al., 2013), mainly due to the absence of such use in all three of the students' first languages.



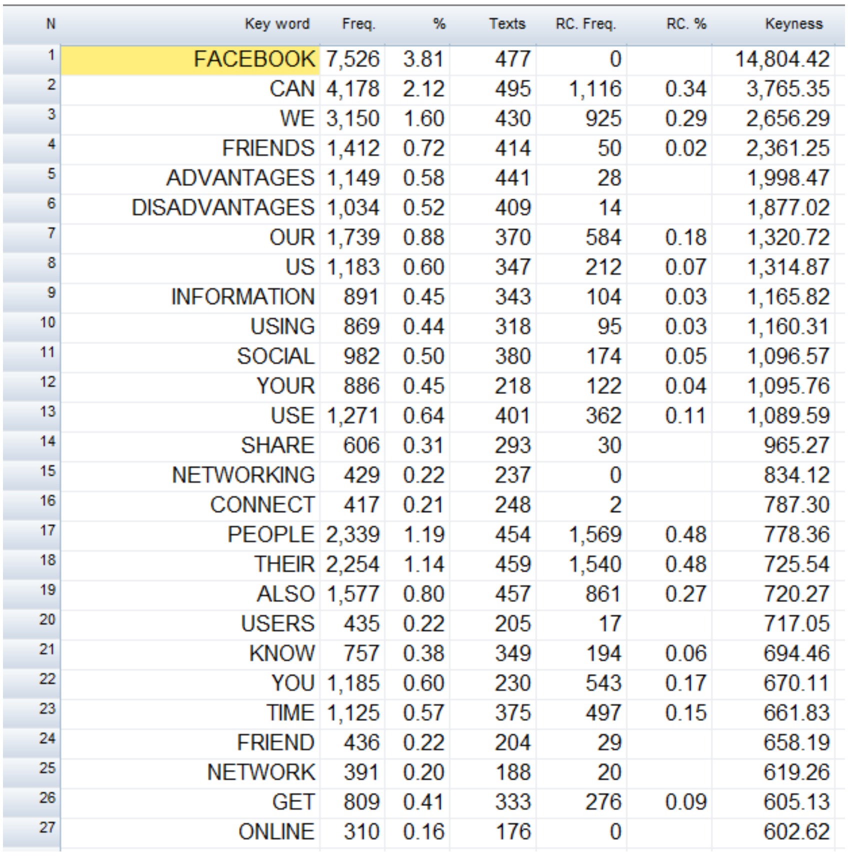
N	File	Words	Hits	per 1,000	Dispersion	Plot
1	FACEBOOK (Overall)	183,87	7,526	40.93	0.923	
2	FACEBOOK 430	1,415	47	33.22	0.874	
3	FACEBOOK 11	1,075	58	53.95	0.869	
4	FACEBOOK 90	677	17	25.11	0.778	
5	FACEBOOK 220	644	20	31.06	0.732	
6	FACEBOOK 381	640	19	29.69	0.746	
7	FACEBOOK 179	624	19	30.45	0.831	
8	FACEBOOK 221	524	24	45.80	0.835	
9	FACEBOOK 319	516	25	48.45	0.772	
10	FACEBOOK 256	514	10	19.46	0.648	
11	FACEBOOK 183	509	30	58.94	0.883	
12	FACEBOOK 315	502	25	49.80	0.791	
13	FACEBOOK 339	492	24	48.78	0.852	
14	FACEBOOK 173	491	27	54.99	0.754	
15	FACEBOOK 308	487	26	53.39	0.787	
16	FACEBOOK 297	484	12	24.79	0.553	
17	FACEBOOK 24	476	24	50.42	0.865	
18	FACEBOOK 210	474	7	14.77	0.723	
19	FACEBOOK 202	472	18	38.14	0.767	
20	FACEBOOK 12	468	23	49.15	0.916	
21	FACEBOOK 311	467	17	36.40	0.639	
22	FACEBOOK 174	466	32	68.67	0.774	
23	FACEBOOK 190	466	19	40.77	0.660	
24	FACEBOOK 320	460	23	50.00	0.890	
25	FACEBOOK 348	460	17	36.96	0.693	
26	FACEBOOK 299	456	23	50.44	0.836	
27	FACEBOOK 205	439	27	61.50	0.831	

Figure 3.1: Sample of plot function for *Facebook* in MCSAW

In the context of the present study, analysing dispersion enables us to examine where words are mentioned in particular stages of the argumentative essay. For instance, Figure 3.1 shows that *Facebook* is highly scattered throughout the texts in MCSAW. This is explainable due to the topic of 'Facebook', which is one of the two essay topics constituting MCSAW. Scattering of the plot thus indicates where an item occurs and, in turn, indicates something about the ways (more specifically the areas) in which it is used in the text.

As mentioned earlier, a corpus-driven investigation often starts with word lists or lists of keywords. WordSmith's KeyWords tool compares word lists (like the one shown in Table 3.2) from two corpora to generate a keywords list, with the help of tests of statistical significance (McEnery & Hardie, 2012: p. 51). Table 3.3 shows an extract of the keywords list taken from MCSAW, with keywords presented both in terms of raw frequency and in terms of their relative frequencies (RC. Freq).

Table 3.3: Sample of MCSAW keywords list



N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness
1	FACEBOOK	7,526	3.81	477	0		14,804.42
2	CAN	4,178	2.12	495	1,116	0.34	3,765.35
3	WE	3,150	1.60	430	925	0.29	2,656.29
4	FRIENDS	1,412	0.72	414	50	0.02	2,361.25
5	ADVANTAGES	1,149	0.58	441	28		1,998.47
6	DISADVANTAGES	1,034	0.52	409	14		1,877.02
7	OUR	1,739	0.88	370	584	0.18	1,320.72
8	US	1,183	0.60	347	212	0.07	1,314.87
9	INFORMATION	891	0.45	343	104	0.03	1,165.82
10	USING	869	0.44	318	95	0.03	1,160.31
11	SOCIAL	982	0.50	380	174	0.05	1,096.57
12	YOUR	886	0.45	218	122	0.04	1,095.76
13	USE	1,271	0.64	401	362	0.11	1,089.59
14	SHARE	606	0.31	293	30		965.27
15	NETWORKING	429	0.22	237	0		834.12
16	CONNECT	417	0.21	248	2		787.30
17	PEOPLE	2,339	1.19	454	1,569	0.48	778.36
18	THEIR	2,254	1.14	459	1,540	0.48	725.54
19	ALSO	1,577	0.80	457	861	0.27	720.27
20	USERS	435	0.22	205	17		717.05
21	KNOW	757	0.38	349	194	0.06	694.46
22	YOU	1,185	0.60	230	543	0.17	670.11
23	TIME	1,125	0.57	375	497	0.15	661.83
24	FRIEND	436	0.22	204	29		658.19
25	NETWORK	391	0.20	188	20		619.26
26	GET	809	0.41	333	276	0.09	605.13
27	ONLINE	310	0.16	176	0		602.62

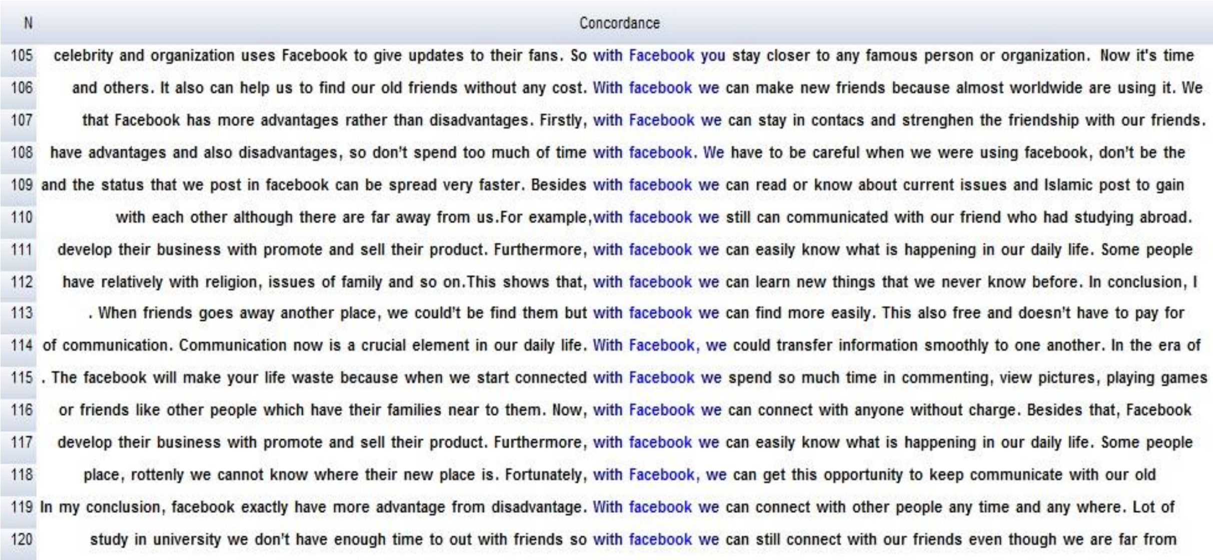
The keywords list is different to a word list because it is the result of a comparison of two wordlists, and features saliency rather than frequencies per se. In other words, a keywords list presents words that appear unusually more frequently or infrequently in one corpus than in the other (reference corpus) (see also Section 3.4.2). As a result, different reference corpora will produce different keywords. Since the tool works by comparing word lists, it can be applied to both lists of individual word forms and to lists of lexical bundles. Hence, the tool can be used to produce key *bundles* (i.e. combinations of words that are statistically more or less frequent) in a similar way to key*words*, although the actual procedure is slightly more complex for the user.[34] As regards the contrastive corpus-driven study, an examination of keywords and key

[34] Bundles are computed automatically by WordSmith Tools (after an index has been compiled) and by imposing several restrictions: choosing how many words a bundle should have (e.g. 2-word bundle, 3-word bundle), how many of each bundle must be found in the corpus (minimum frequency), and by instructing the tool to stop counting bundles at sentence breaks because “a [bundle] which spans across two sentences is not likely to make sense” (Scott, 2015).

bundles reveals lexico-grammatical patterns between two corpora, illustrating what the essays are about (i.e. content) and how they are written (i.e. writing style).

In addition, the list is usually presented in order of keyness (statistical significance), with the most statistically significant or 'strongest' keywords appearing first. Statistical significance tests "allow researchers to assert with a degree of confidence that the results of their analysis either are or are not significant" (McEnery & Hardie, 2012: p. 51). The statistical operations involved, i.e. "a cross tabulation and a chi-square or log likelihood significance test, are basic and commonly used in corpus linguistics" (Culpeper, 2009: p. 34). In the present thesis, log-likelihood was used, following arguments made by McEnery and Hardie (2012) that preference for this test by some corpus linguists is because "it makes no assumption of a normal distribution" (McEnery & Hardie, 2012: p. 52) compared to the chi-square.[35]

Also, included in WordSmith is the Concord Tool. One common use of this tool is in examining the occurrence of words in their respective textual environments, i.e. Key Word in Context (KWIC). An example of a KWIC concordance of the word *Facebook* in MCSAW is shown in Figure 3.2.



N	Concordance
105	celebrity and organization uses Facebook to give updates to their fans. So with Facebook you stay closer to any famous person or organization. Now it's time
106	and others. It also can help us to find our old friends without any cost. With facebook we can make new friends because almost worldwide are using it. We
107	that Facebook has more advantages rather than disadvantages. Firstly, with Facebook we can stay in contacs and strenghen the friendship with our friends.
108	have advantages and also disadvantages, so don't spend too much of time with facebook. We have to be careful when we were using facebook, don't be the
109	and the status that we post in facebook can be spread very faster. Besides with facebook we can read or know about current issues and Islamic post to gain
110	with each other although there are far away from us.For example,with facebook we still can communicated with our friend who had studying abroad.
111	develop their business with promote and sell their product. Furthermore, with facebook we can easily know what is happening in our daily life. Some people
112	have relatively with religion, issues of family and so on.This shows that, with facebook we can learn new things that we never know before. In conclusion, I
113	. When friends goes away another place, we could't be find them but with facebook we can find more easily. This also free and doesn't have to pay for
114	of communication. Communication now is a crucial element in our daily life. With Facebook, we could transfer information smoothly to one another. In the era of
115	. The facebook will make your life waste because when we start connected with Facebook we spend so much time in commenting, view pictures, playing games
116	or friends like other people which have their families near to them. Now, with Facebook we can connect with anyone without charge. Besides that, Facebook
117	develop their business with promote and sell their product. Furthermore, with facebook we can easily know what is happening in our daily life. Some people
118	place, rottenly we cannot know where their new place is. Fortunately, with Facebook, we can get this opportunity to keep communicate with our old
119	In my conclusion, facebook exactly have more advantage from disadvantage. With facebook we can connect with other people any time and any where. Lot of
120	study in university we don't have enough time to out with friends so with facebook we can still connect with our friends even though we are far from

Figure 3.2: Concordance lines for *Facebook* in MCSAW

[35] "Data has a normal distribution if most of the values cluster relatively tightly around a mean (average) value – a pattern which, when plotted on a graph, gives us the classic 'bell-shaped' curve. This is not true for language data" (McEnery & Hardie, 2012: p. 51).

WordSmith’s Concord tool locates words in a corpus and shows them in standard concordance lines: the search word centred with a variable amount of context (surrounding text) at either side (Scott, 2001; Adolphs & Lin, 2011). The concordance lines, as illustrated in Figure 3.2, allow for further examination of the company a given word keeps, i.e. the words that surround the word or phrase of interest. *Facebook* (in MCSAW), as shown in the figure above, gives us a sense of the word being used in the text and its possible patterns in learner language use. For instance, *Facebook* is seen to co-occur quite frequently after the preposition *with* in the phrase ‘with Facebook’, appearing mostly at the beginning of a sentence. This allows for a practical linguistic analysis of a word to be conducted, especially in investigating its meanings and discourse functions. In addition to concordance lines, the Concord tool also provides the user with information about word forms that repeatedly co-occur (collocation), which is explained in Section 3.4.3 below.

This sub-section has shown general uses of the corpus software employed in the present study. As observed in LCR scholarship, WordSmith Tools is commonly implemented in many studies (Breeze, 2007; Gilquin & Paquot, 2008; Römer, 2009). This is mainly because WordSmith Tools has plenty of features for analysing language. It appears to be particularly efficient at handling large amounts of data quickly, giving prompt and detailed statistics, and enabling swift movement from one function to the other, and it is very flexible for the users’ needs. For the purpose of this thesis, WordSmith Tools is found to be useful, and is therefore employed for data analysis. The following section will discuss in more detail how the suite of tools is applied in the thesis.

3.4 Step by step procedure

Following Marchi (2013: pp. 86-87), the analytical process in this thesis can be graphically described as the process of ‘funnelling’ (see Figure 3.3), with the exception that analysis in the present study starts with analysing keywords in MCSAW (compared to analysing wordlists, as in Marchi 2013). Collocation analysis is then used in both Chapter 4 and Chapter 5 to explore the co-occurring words surrounding the keywords *can* and *we*, for a view of some typical patterns in learner writing. As explained in Chapter 2, collocation refers to the non-random co-occurrence of words in a corpus, with a particular node word typically having a range of collocates which are automatically determined by the software. Concordance analysis is then used to examine individual keywords (*can* and *we*) and key bundles in more detail, which shows the use of these words and phrases in context (in Chapters 4, 5 and 6, respectively). Finally, comparison and evaluation of both corpora are discussed in all analysis chapters.

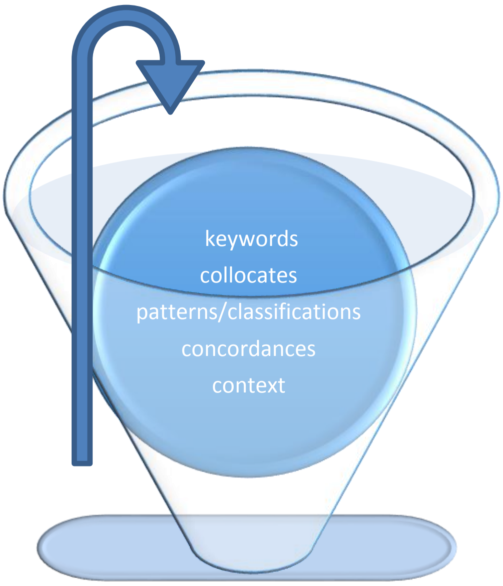


Figure 3.3: Graphic representation of the analytic process, modified (Marchi, 2013)

The following sub-sections offer more detailed explanations of the step-by-step process illustrated in Figure 3.3. This is in order to present the methods more clearly, while at the same time providing systematic documentation of the steps taken. The following discussion is divided into five parts: selecting items (3.4.1); analysing keywords (3.4.2); analysing collocates (3.4.3); analysing concordance lines (3.4.4); and comparing and contrasting (3.4.5).

3.4.1 Selecting items for a corpus-driven approach

The first thing to do after extracting keywords from MCSAW is to scan for interesting items to be analysed further. In doing so, three criteria are borne in mind: items should be useful in that they help answer the research questions; items should be valid according to the approach taken; and items should be practical in the sense of their use (Marchi, 2013: p. 88). Like many corpus studies, the data is firstly examined in terms of their relevance to the research questions and then in terms of comparability in conducting a CIA approach. To reinstate, the RQs are:

1) What are the most salient linguistic items found in the Malaysian learner corpus compared to those in the reference corpus?

2) How are the items used similarly or differently in the two corpora (including their collocations)?

3) What are the most overused types of lexical bundles found in MCSAW?

4) How do these bundles function in Malaysian learner argumentative writing?

In this thesis, an item is useful if it is related to the description of learner language in general, and to the writing style in particular. By using the keywords technique to compare Malaysian learner writing with the reference corpus, it is possible to identify a series of statistically significant words in MCSAW, to categorise them, and to examine their range. This preliminary analysis corroborates previous research (Luzón, 2009; McCrostie, 2008; Mukundan et al., 2013), revealing that modality and personal pronouns were among the most characteristic (also problematic) of learner language. These two main findings are analysed into two separate chapters: the highly frequent use of *can* is considered in Chapter 4, where the use of the modal verb is investigated in terms of its polysemous meanings compared to its use in LOCNESS; I then continue to analyse the highly frequent personal plural pronoun *we* in Chapter 5 in terms of its discourse functions, along with comparisons made to its use in LOCNESS. Consequently, these two chapters seek to answer RQs 1 and 2.

In response to the second criterion, validity is ensured by the means that words investigated via the corpus-driven approach should "emerge bottom-up from the data, rather than being selected intuitively (or introspectively)" (Marchi, 2013: p. 88). In other words, words are selected through extracting keywords, using statistical measures. This, according to Bondi (2010: p. 3), can "point to elements that may be profitably studied and need to be explained". As Lee and Chen (2009: p. 153) further highlight, the corpus-driven method thus "differs from the more deductive approach of predetermining a number of words that might be problematic on the basis of linguistic intuition or teaching experience, then going to a corpus to find the instances and trying to account for them". As previously described, keywords list functions as a starting point in which items are extracted from the comparison of the target and reference corpora.

Finally, selecting items that are practical excludes those that are highly topical (e.g. Facebook), or when referring to a proper pronoun (e.g. Malaysia). These would not be interesting, as they are too specific to only one corpus. In addition, some word forms are not explored further due to their polysemy. In spite of the exclusion of some items, investigation of both individual and recurrent word combinations (i.e. bundles) (Section 3.4.2) is conducted, for the extent to which they allow us to answer the questions we wish to ask, specifically in terms of lexis and phraseology. Lexical bundles are therefore the focus of Chapter 6, where the use of different types of lexical bundles in MCSAW is explored, following the scholarship in recurrent word combinations in learner writing. This is then compared to how bundles are used in LOCNESS, answering RQs 3 and 4, concurrently.

3.4.2 Analysing keywords and key keywords

An analysis of keywords is often the first step in contrastive corpus-driven research. As a reminder, keywords are words that occur significantly higher or significantly lower in a text or collection of texts, when compared to a reference corpus. Keywords are thus a good starting point because "a keyword [is] a word that is statistically characteristic of a text or texts" (Culpeper, 2009: p. 30). In investigating keywords, Culpeper (2009) argues for three questions to be considered:

1) What decisions need to be made in performing a keyword analysis?
2) What kinds of keywords result from an analysis?
3) Are all keywords general features of the data in focus?

Before one conducts a keywords analysis, the choice of data for comparison (i.e. the reference corpus) is critical, because it will influence the keywords revealed (Culpeper, 2009: p. 34). Culpeper further argues that "[t]he closer the relationship between the target corpus and the reference corpus, the more likely the resultant keywords will reflect something specific to the target corpus" (2009: p. 35). In this thesis, argumentative essays in MCSAW are compared with argumentative essays in LOCNESS, which will reveal differences between how the two groups of novice writers construct their essays, and keywords that are particularly more distinctive of MCSAW writers. However, a caveat lies in the differences of topic presented in the corpora, which is likely to result in highly significant topic-related words.

Other considerations include the minimum frequency for a word to be considered key, and the test for statistical significance. As mentioned earlier in Section 3.3.4, the log likelihood test was used to calculate key words. Following Culpeper (2009), due to the relatively small data set in the thesis, the minimum frequency for a word to be considered key was set at five, and the probability value (i.e. p-value) was set to smaller than or equal to 0.01. This means that words are considered keywords if their differences were considered to have a 1% chance or less of being coincidental (i.e. to happen by chance). As regards the token definition (determining what counts as one 'word'), contractions such as *don't* were counted as one word by the software. In addition, items that contain hyphens (e.g. *notice-board*) and numbers[36] (e.g. *$1000*) are also considered as one word.

Once keywords are extracted, there is a distinction between positive keywords (unusually frequent in the target corpus) and negative keywords (unusually infrequent in the target corpus). For the purpose of the present study, positive keywords are found to be more interesting to be investigated as they are words that are more characteristic of Malaysian learner writing. It is also common in corpus linguistics to focus more on 'positive' (overused) rather than 'negative' (underused) keywords. It has been found that there are three kinds of keywords: proper nouns; (lexical) keywords that relate to 'aboutness' or content; and (non-lexical) indicators that are more of style than aboutness (Culpeper, 2009; Scott & Tribble, 2006).[37] In Chapter 4, individual keywords are firstly categorised based on these three kinds of keywords. This includes further classifying types of functional keywords and examining their (key) keyness values. The classification of keywords can be seen as advantageous, as Bondi (2010: p. 3) states, in that it "point[s] to fundamental elements in describing specialised discourse" as well as systematically "placing a text in a specific domain".

Finally, Culpeper warns that "it is easy to retrieve keywords that are key, but not actually general features of the data one is examining" (Culpeper, 2009: p. 39). He argues that this results in some highly misleading characterisations of particular discourses or genres. Consequently, Culpeper (2009), along with many others (e.g. Baker, 2004; Rayson, 2008; Gries, 2013), advocates for examining the distribution of keywords, i.e. whether they are localised or well-distributed throughout the corpus (see further in Chapter 4). If this is not taken into account, then descriptive results are fairly random (Brezina & Meyerhoff, 2014).

[36] Numbers that are ignored (i.e. not counted as words) in word lists, key words, concordances etc., are replaced by a # (Scott, 2015).

[37] This is described in further detail, along with the analysis, in Chapter 4.

Generally, in this thesis, three main steps are taken in analysing keywords. As previously explained, ‘range’ is used as a down-sampling technique, in which keywords are selected not only by their statistical significance but also distribution across the corpus. In addition, I also looked at L1 backgrounds, and analysed how proportionate keywords are used across different learner groups in MCSAW. I also used the dispersion plot to investigate the textual position of keywords.

Another procedure to check the distribution of keywords is to make use of WordSmith’s extraction of ‘key-keywords’ (Scott, 2012). Such words are not more key than other keywords, but are words that are keywords in a number of different files. In other words, ‘key-keywords’ are considered “keywords […] that are well-dispersed across many different texts of the study [or target] corpus rather than clumped in just a few idiosyncratic ones” (Lee & Chen, 2009: p. 153). By examining ‘key-keywords’, we can be more confident that the keywords we select are words that are generally key across the body of data as well as general features of MCSAW in particular. While my starting point is a standard keywords analysis, I will test whether the keywords I select for analysis are also ‘key-keywords’.

Finally, as noted above, this thesis also investigates key bundles (Chapter 6), which are categorised based on frameworks proposed for different functions of bundles, distinguishing between ‘referential’, ‘discourse-organising’, and ‘stance bundles’ (Biber et al, 2004; Chen & Baker, 2014). In terms of the settings for producing key bundles, I follow Biber et al. (2004) and Cortes (2004) in investigating 4-word lexical bundles, which are argued to be more frequent than 5-word bundles. 3-word bundles are also investigated in order to distinguish whether they make up longer 4-word strings of words, as found in Cortes (2004: p. 401). Similar to the software settings for keywords, contractions are counted as one word (i.e. *we don’t* would be identified as bigram), and words that consist of hyphens are not counted as separate (i.e. *the notice-board* would be considered as bigram). In addition, bundles are not calculated across sentence boundaries as this is not likely to make sense (Scott, 2015). The minimum frequency for a bundle to be included was again set at five, and the p-value set to smaller than or equal to 0.01.

As will be shown in Chapter 6, mainly three types of procedures are applied when analysing key bundles. Firstly, like the individual keywords analysed in Chapters 4 and 5, key bundles are extracted the same way, but based on lists of bundles rather than on word forms (see Footnote 32 in Section 3.3.4). Selection of key bundles was determined on the basis of at least one occurrence in the reference corpus. Following this, key bundles that occur in both corpora are classified according to discourse functions. Finally, instances of different types of bundles are examined further through investigating concordance lines. Findings regarding lexical bundles will be compared with those of other researchers, where relevant, and in some cases with data from COCA (Corpus of Contemporary American English). In sum, as depicted in Figure 3.3, the thesis begins with keywords analysis to identify salient words (and salient lexical bundles) that are more significant in MCSAW compared to in LOCNESS, and continues the corpus-driven approach with collocational and concordancing, which are described next.

3.4.3 Analysing collocates

As mentioned in Chapter 2, the approach to collocation analysis in this thesis is statistical, meaning that collocates are determined through statistical association measures commonly used in corpus linguistics. According to McEnery and Hardie (2012: p. 51), “we test the significance of the co-occurrence frequency of [a] word and everything that appears near it once or more in the corpus” when identifying a word’s collocates. The two most commonly used measures in collocation analysis are the t-score and Mutual Information (MI) value (Cheng, 2012; Hunston, 2002b). These measures are used to determine whether two words co-occur by chance, or whether they are co-selected by the speaker/writer, in which latter case their association becomes significant. As Hunston (2002b: p. 70) notes, “t-score uses a calculation of standard deviation” and gives measures for evidence or confidence that words are associated with each other, usually indicating significance if scores are 2 or higher (Hunston, 2002b: p. 72). MI, on the other hand, measures the collocational strength of the connection between the node word and each collocate, and “[a]n MI-score of 3 or higher can be taken to be significant” (Hunston, 2002b: p. 71). Baker explains the MI score as follows:

> Put simply, mutual information is calculated by examining all of the places where two potential collocates occur in a text or corpus. An algorithm then computes what the expected probability of these two words occurring near to each other would be, based on their relative frequencies and the overall size of the corpus. It then compares this expected figure to the observed figure - what has actually happened, and converts the difference between the two into a number which indicates the strength of the collocation - the higher the number, the stronger the collocation (Baker, 2006: p. 101).

However, it has been argued that MI gives too much prominence to rare combinations (Lindquist, 2009: p. 76), whereas the t-score is more likely to extract highly frequent words (often privileging function words) than one based on MI (Cheng, 2012; Hunston, 2002b). Consequently, throughout the analysis in this thesis, collocates are derived from both MI and t-score, with cut-off points decided at 3 and 2, respectively - calculated using WordSmith Tools. This means that a combination of both statistically significant function words and strongly associated lexical words can be explored.

In terms of the collocational span and setting of thresholds, corpus linguists and computational linguists usually work with either a span of +/- 4 or +/-5 (Brezina, et al., 2015: p. 140). In this thesis, I investigate collocates within a span of five words to the left and right throughout the analysis. Given the size of the comparable corpora used, I set the minimum co-occurrence frequency to 3. Hence, within a 5:5 window span, items which have a minimum co-occurrence frequency of 3 as a collocate of a given node word, and a minimum t-score of 2 and MI score of 3, are considered to be collocates of a node word. In this thesis, collocation analysis is used to investigate collocates of two highly significant items (*can, we*) in MCSAW and LOCNESS, and whether they are common to both language varieties or not.

3.4.4 Analysing concordance lines

To reiterate, a concordance is "a collection of the occurrences of a word-form, each in its own textual environment" (Sinclair, 1991: p. 32). In contrast to frequency lists, a concordance analysis combines quantitative and qualitative analysis by allowing researchers to carry out close examination of a word in context (Hunston, 2002b: p. 129).

As shown in Section 3.3.4 earlier, concordances are read vertically, where the target item, i.e. the node word, is positioned in the middle of surrounding context words. Repeated co-occurrences of the node with other words or phrases can emerge from the concordance and can take the shape of a pattern. This is carried out by examining each key item (i.e. *can*, *we* and key bundles) that have been selected for further qualitative analysis. Analysis of discourse functions are also conducted for keywords in Chapters 4, 5, and 6, in order to identify how they are used. As will be described in the subsequent chapters, classifications of items according to their discourse functions are carried out based on respective research, namely on modal meanings (Coates, 1983; Palmer, 1990, 2001), discourse functions of personal pronouns

(Luzón, 2009), and on use of lexical bundles (Bednarek, 2008a; Biber et al., 2004; Chen & Baker, 2014; Chen & Baker, 2010).

In Chapter 4 and Chapter 5, the analyses are mainly focused, in particular, on patterns of the way the modal verb *can* and personal pronoun *we* are used in Malaysian learner English argumentative writing. After exploring the collocational patterns of the items under investigation, concordance lines are investigated with reference to types of meanings described in the literature, in order to make close reading of the context in which the keywords are situated in texts. In Chapter 6, however, I do not examine collocates of key lexical bundles; instead, I examine their functional categories alongside close examinations of concordance lines for contextualisation purposes. Concordance analysis, therefore, is essential for interpreting patterns in a larger context.

3.4.5 Compare and contrasting

As repeatedly mentioned, the CIA approach is comparative in nature (2.3.3), and the literature in LCR highly advocates this approach (Granger, 1992; 2015). Throughout the investigation of Malaysian learner English writing via the contrastive corpus-driven approach, in this thesis, there are three various ways in which the two language varieties (MCSAW and LOCNESS) are compared:

- Comparison of words and bundles that are more frequent, statistically speaking, in MCSAW than in LOCNESS. This is done by analysing keywords and key bundles across portions of text;
- Collocational comparisons of the two individual keywords (*can* and *we*) across both corpora, and comparing collocates for their strength or statistical significance;
- Comparison of discourse functions of selected keywords and key bundles. More specifically, comparing how the same word or phrase is used in the different varieties.

Although findings are discerned from the contrastive corpus-driven approach, it is important to bear in mind that results are only "relative to the comparison made and that a different operationalization will most likely produce different results" (Marchi, 2013: p. 101). This is particularly evident with keywords analysis, as discussed in Section 3.4.2. Therefore, it is important to be cautious when interpreting data, and to reflect on the premise that there are limitations with every contrastive type of corpus study, as discussed in Chapter 7. Nevertheless, the steps carried out in this study have attempted to remain as close to the CIA approach as possible. Where comparability in data is concerned, the investigation of MCSAW and LOCNESS is suitable given the genre of argumentative writing as well as participants of the corpora being novice writers. It is also hoped that the contrasting features of certain procedures in the thesis have been described in detail, providing justifications to some of the decisions taken, as well as narrating as closely to the actual research process as possible.

3.5 Summary

This chapter presented how comparison between Malaysian argumentative writing and a reference language variety is conducted in this thesis via a contrastive corpus-driven approach. This includes how features of the learner language were identified empirically, rather than introspectively, and how the analysis itself reflects Marchi's (2013) model of recursive funnelling (Figure 3.3). The analysis starts off from quantitative observations (through keywords list), followed by qualitative interpretations (via concordancing) – of both individual lexical items and lexical bundles – through bottom-up handling of the data. This thesis also highlights three specific procedures that enhance the qualitative analysis of corpus methods, namely by looking at how keywords are used across different groups of Malaysian writers, by using range to identify and confirm saliency of keywords that are distributed widely across texts (i.e. across learners), and by analysing a dispersion plot that illustrates the scattering of keywords in particular sections of a text (i.e. generic stages).

In the following three chapters, observations of Malaysian learner English writing and the reference language variety are evaluated and analysed, each chapter providing answers to the research questions. Chapter 4 focuses on general keywords, and then investigates how modality (specifically, *can*) is used in learner writing, and what distinctions are found among novice writers of MCSAW and LOCNESS. Chapter 5 then examines how writer visibility impacts both MCSAW and LOCNESS novice writing, by analysing the personal pronoun *we*. Chapter 6, finally reports on how and to what extent Malaysian learners use lexical bundles in comparison to the reference language variety.

Chapter 4: Keywords and the Modal Verb *can*

4.1 Introduction

The first step of conducting a corpus-driven analysis is by examining a keywords list. As discussed in Chapter 3, a keywords list shows the most statistically significant words occurring in a target corpus relative to its occurrence in a reference corpus. Essentially, keywords illustrate what the collection of texts is generally about and hence, highlight words that are more salient for further investigation. This chapter presents results for keyword analysis of the Malaysian corpus (MCSAW) against its reference language variety, LOCNESS via Wordsmith Tools Version 6 (Scott, 2012) as outlined in Chapter 3. Keywords are categorised into two broad types, namely functional and lexical keywords in Section 4.1.1. These keywords are then discussed in terms of their keyness value, range and collocates, which reveal significant insight into the selection of keywords to be analysed. Two most significant findings include patterns of modality (with emphasis on the modal verb *can*) and pronouns (specifically *we*) used differently by Malaysian learners than their native speaker counterparts. In Section 4.2, *can* is given emphasis, drawing on past scholarship of modality and using other corpus methods for further analysis, followed by a close examination of its meanings as used in MCSAW and LOCNESS in Section 4.3. Pronouns will be discussed in the subsequent chapter.

4.1.1 Key words

As noted in Chapter 3, keywords are words whose frequency (or infrequency) in a text or corpus is statistically significant and therefore, worth investigating. Research has asserted that examining how keywords occur in context and which grammatical categories they appear in, and looking at their common patterns of co-occurrence can be revealing. Scott (1999), cited in Baker (2004: p. 347), emphasises that three types of keywords are usually found:

> proper nouns; keywords that human beings would recognize as key and are indicators of the "aboutness" of a particular text; and finally, high-frequency words such as *because, shall* or *already*, which may be indicators of style, rather than aboutness.

Accordingly, Table 4.1 shows all the keywords in MCSAW (i.e. both positive/overused and negative/underused) further classified into these types, namely functional, lexical, and proper noun categories. [38] It is important to highlight that this thesis focusses primarily on overused keywords, while underused keywords constitute an important area for future research. Consequently, I will only occasionally comment on negative keywords.

Overall, the seven proper nouns show a sense of what majority of texts in MCSAW is about and therefore, it is unsurprising that *Facebook* is identified as the top keyword (see Table A4.1 in Appendix) in the Malaysian corpus. The explanation for this lies in the nature of one of the two essay topics, which learners are required to write about, that is 'Do you think Facebook has more advantages than disadvantages? Discuss your reasons'. The proper nouns therefore, signal words that are often used in association with social media, i.e. *Facebook*, *Twitter*, *Zuckerberg*, *Mark*, *Facebook's*, and *Yahoo* in response to this particular essay prompt. In the reference language variety (henceforth, LOCNESS), essay topics are various and none of these topics are related to Facebook. With the exception of *Malaysia* then, all of the proper nouns in MCSAW are clearly topic-related, and will not be discussed any further.

There are more lexical keywords than functional keywords in Table 4.1, which is typical of a keywords list. As mentioned by Scott and Tribble (2006: p. 63), "keywords are mostly connected to what the text is about and are important to it, with some intruders which suggest something about the style and which often repay further analysis". These "intruders" refer to grammatical words such as *can*, *we*, and *with* that are suggestive of the writing style (as discussed below), whereas keywords that are connected to the aboutness of the text are indicative of the topics in MCSAW. Closer observation reveals that a number of these lexical keywords like *networking*, *connect*, *online*, *internet*, *chat* and *website* describe communication and technology, which are mostly related to the topic 'Facebook'. Keywords that refer to student accommodation such as *hostel*, *stay*, *expenses*, and *rent* on the other hand, are closely associated to the topic 'Living in a Hostel'. Hence, like the proper nouns, many lexical keywords identified in MCSAW are also topic-based and for the purpose of this chapter, will not be analysed in more detail.

[38] Table A4.1 in Appendix contains all keywords (both positive/overused and negative/underused) with raw frequencies, range, and relative frequencies, arranged according to keyness values (with the most statistically significant, i.e. "strongest" keywords appearing first).

Table 4.1: Categories of keywords[39]

Functional (grammatical)	**Positive**: about, addition,[40] all, almost, also, among, and, any, anything, anytime, anywhere, apart, around, be, because, beside, besides, can, do, don't, everyone, firstly, for, foremost, from, furthermore, have, it, it's, lastly, like, lot, many, more, moreover, most, need, nowadays, or, other, others, our, secondly, so, some, someone, sometimes, than, their, them, thirdly, through, too, us, via, we, will, with, without, you, your
	Negative: a, already, an, at, be, been, being, before, could, did, does, down, ever, had, he, her, his, however, if, into, less, may, no, not, of, only, out, over, she, should, that, the, themselves, these, this, those, two, was, were, what, which, would
Lexical	**Positive**: abroad, account, activities, actually,[41] add, addicted, addicting, addiction, advantage, advantages, advertise, advertisement, advertising, agree, application, applications, assignment, assignments, avoid, bad, become, benefit, benefits, best, biggest, billion, block, brings, browsing, business, busy, button, call, careful, carefully, chat, chatting, click, colleague, comment, communicate, communication, conclusion, connect, connected, connecting, connection, cons, contact, convenient, cost, country, create, creating, culture, custom, customers, daily, date, depend, depends, disadvantage, disadvantages, discuss, discussion, drastically, easier, easily, easy, era, especially, example, exams, expenses, face, fake, family, famous, fan, faster, features, feedback, feelings, find, finding, fine, free, friend, friendship, front, gain, games, gather, get, give, gives, good, group, groups, harass, harm, harming, help, helping, helps, homework, hostel, id, important, income, information, instance, insult, interact, internet, keep, know, knowledge, laptop, largest, latest, like, limit, log, lost, low, make, manage, marketing, marks, medium, meet, message, messages, minimize, network, networking, networks, new, news, nutshell, offices, old, online, opinion, opportunity, page, pages, people, personal, phone, photo, photos, picture, pictures, place, platform, playing, popular, popularity, post, precious, priority, privacy, private, product, products, profile, projects, promote, proper, properly, pros, publicly, relationship, relatives, rent, save, search, send, share, sharing, site, sites, smart, social, spend, spread, stalk, status, stay, strongly, student, students, studies, study, studying, technology, teenagers, tend, things, thoughts, time, tool, touch, tradition, trouble, update, updated, updates, upload, use, useful, user, users, using, valuable, video, videos, wall, want, waste, wastes, wasting, ways, website, wisely, world
	Negative: believe, better, case, cases, change, children, community, fact, feel, human, issue, job, lives, made, number, point, problems, public, say, seen, single, society, suicide, times, years
Proper nouns	**Positive:** Facebook, Facebook's, Malaysia, Mark, Twitter, Yahoo, Zuckerberg **Negative: -**

[39] In this table, nouns, adjectives, verbs and adverbs have been categorised as lexical whereas prepositions, conjunctions, and pronouns have been classified as functional (grammatical). Functional keywords also include modal verbs, numerals, and adverbs that are not derived from adjectives. Verbs *do, be, have* and their forms are also categorised as functional keywords although they can act as both auxiliary and linking verbs. Proper nouns are categorised separately. It is worth pointing out that there is ambiguity of grammatical categories, and in turn, ambiguous words such as *like* are double-classified under both word forms (grammatical/functional and lexical). In general, many words in English belong to different categories and it is then not possible to establish which category a word belongs to out of context. This also affects the sub-categorisation of functional words in Table 4.2, which also contains words that are double-classified.

[40] Classified as 'functional' because of likely use as conjunction/discourse marker (*in addition*)

[41] And *especially* classified as 'lexical' because of adjectives *actual* and *special*, but could also be considered as 'grammatical'

In contrast, functional keywords are more interesting to explore because they indicate the structure or style of writing in a particular text other than its contents. This means that the investigation of grammatical words such as *can*, *we*, *us, our*, *your,* and *also* can show us the ways in which MCSAW speakers write their essays. Functional or grammatical words are mostly studied in learner corpus research for investigating different or idiosyncratic uses of words in grammar (e.g. Granger & Tyson, 1996 for the overuse of connectors among French learners) or describing patterns of local grammar which help explain learners' language development (Hunston, 2002a; Nesselhauf, 2003). Essentially, these types of keywords are imperative to study learners' language proficiency at the same time useful for contrastive analyses.

Like the complete keywords list, functional keywords can be further categorised in terms of part of speech, i.e. determiners, prepositions, pronouns, conjunctions, modals or others, as shown in Table 4.2. Functional keywords were classified by consulting Quirk & Greenbaum (1975), and Halliday & Hasan (1976), with ambiguous cases classified more than once. For example, *like* can be a preposition (e.g. *She looks **like** her mother*) or a conjunction (e.g. *Nobody understands her **like** I do*), so has been classified twice. Conjunctions are further classified into six headings, namely addition or additive (used to signal addition, introduction, similarity, etc.); comparison or adversative (used to signal conflict, concession, etc.); time or temporal (in relation to both temporal and textual times); cause or causal (used to signal cause/effect, reason/result, etc.); words indicating condition; and purpose (used to signal a chronological or logical sequence). These categories have been classified according to Halliday and Hasan (1976) as well as discussion in past research on learners' use of cohesive devices in writing (Liu & Braine, 2005; Yang & Sun, 2012). It is important to note that 'time' also relates to the marking of textual time (discourse structure), i.e. *firstly, secondly, lastly* etc.

Table 4.2: Further classification of functional keywords

Category	**Keywords**	
	POSITIVE	**NEGATIVE**
Article/quantifier/ negator/relative pronoun, etc.	all, almost, any, lot, many, more, most, some	an, a, the, less, no, not, two, what, which, that
Verb	be, can, do, don't, have, it's, need, will	be, been, being, could, did, does, had, may, should, was, were, would
Preposition	about, among, apart, around, beside, from, for, like, via, than, through, with, without	at, before, down, into, of, out, over
Conjunction:		that[43]
Addition	addition, also, and, besides, furthermore, moreover, or	-
Comparison	like, than	however, only
Time	firstly, foremost,[42] lastly, secondly, thirdly	-
Cause	because, for, so	-
Condition	-	if
Purpose	-	-
Pronoun/determiner	anything, everyone, it, other, others, our, someone, their, them, us, we, you, your	he, her, his, she, themselves, these, those, this, that
Adverbs	anytime, anywhere, nowadays, sometimes, too	already, ever

Table 4.2 shows that there are more determiners, prepositions, conjunctions, pronouns and adverbs in MCSAW compared to their use in LOCNESS (positive keywords). Most verbs and modals on the other hand, are under-used (negative keywords). While the common articles *a*, *an*, and *the* are also under-used in the Malaysian corpus, other determiners such as *any, many, more*, *most*, *all*, *some* are widespread in MCSAW. One reason for this relates strongly to the

[42] This is probably used as *first and foremost*.

[43] As a conjunction, *that* is used to introduce different types of clauses (Biber et al., 2002), and is too ambiguous to be sub-classified further here.

description of Malaysian learners' first language (Malay, Chinese, and Tamil) as being exclusive of articles compared to the English language (Mukundan et al., 2012). The overuse of indefinite determiners (e.g. *everyone*, *anything*, *someone*) reveal similar findings to Granger and Rayson's (1998: p. 122) study in which learners minimise personal reference in their writing, at the expense of using indefinite determiners that are characteristic of speech. Malaysian learners also seem to produce more contractions (e.g. *it's*, *don't*) in their writing, with fewer be-verb forms (i.e. *been*, *were*, *was*, *be*). Similarly, the contractions suggest an influence of spoken colloquial language in writing, while it may be argued that fewer be-verb forms highlight the lack of tense and agreement features in learners' English writing.

There is greater use of prepositions in MCSAW than in LOCNESS despite research that report learners who often under-use many prepositions (Gilquin et al., 2007: p. 323). This is probably because prepositions can be used to signify different meanings. Some prepositions indicate expressions of place/direction/movement (*around*, *through*, *beside*, *from*) while others express means/purpose (*with*, *among*, *via*, *for*, *than*, *without*). For example, Flowerdew (1998: p. 547) argues that prepositions *with*, *through*, *from* and *for* can also function as causative devices, but this key function of prepositions is mainly ignored in EAP textbooks. In turn, learners may become unaware of other ways to use prepositions in their argumentation, which leads to an overuse of prepositions for common purposes. Modals however, occur more frequently in LOCNESS with the exception of *can* and *will*, which is found to be pervasive in MCSAW. [44] Research has shown that the modal verb *can* and *will* are especially frequent in spoken conversations and expository prose, respectively (Kennedy, 2002; Mindt, 1995). Of the two modals, only *can* is in the top 50 keywords (see section 4.1.2), and will be further analysed below.

It can be seen that most conjunctions in MCSAW are used more frequently in relation to signalling addition, comparison, time, and cause. In contrast, conjunctions that signal condition and purpose are not found. Also, most of these conjunctions refer to transitions and frame markers such as *besides* and *firstly*, which according to Hyland (2005: p. 125), "indicate relationships between arguments, [that] help structure the local and global organization in the text".[45] It has been found that learners over-use sentence level conjunctions (*however*, *therefore*, *as a result*) and frame markers used to sequence material (*first*, *second*, *lastly*)

[44] It is important to note that here I discuss matters pertaining to relative frequencies and not raw frequencies.

[45] Halliday and Hassan (1976) classify conjunctions into four main categories, namely additive, adversative, causal, and temporal (pp. 242-243).

compared to English native speakers (Hinkel, 2002; Liu, 2008; Liu & Braine, 2005). This appears to be the case for Malaysian learners of English, specifically for frame markers. Arguably, this can be expected, as the corpus samples contain argumentative essays, where students would be expected to make use of discourse markers and linking expressions in order to logically structure their discourse. While some studies have attributed their findings to interference of the first language (Granger & Tyson, 1996), Milton and Tsang (1993) ascribe students' enthusiasm for transitions to over-teaching in Hong Kong schools. Hinkel (2002) referred this as students' attempt at organising information according to the appropriate structure of essays with the prescribed conventions often taught in schools, i.e. topic or transition markers. Similarly, Malaysian students have been exposed to transitions or conjunctions (and frame markers) that are characteristic of process writing, which is a type of essay writing common in the Malaysian classroom pedagogy (Mohamed Ismail et al., 2013).

Further, there seems to be more use of pronouns in the Malaysian corpus, particularly the first person plural pronouns *we*, *us* and *our*. Use of plural pronouns often indicate plural authorship (or sense of communal justification), while use of the second person pronoun *you* suggests that arguments may be directed to readers. According to a number of studies, high use of these personal pronouns implies a high degree of writer visibility and involvement (Neff et al., 2004; Petch-Tyson, 1998). These studies also suggest that the use of the first-person pronoun as a strategic resource requires a high degree of genre awareness, which learners find difficult to do. More specifically, research has shown that first person pronouns are highly problematic for learners, who tend to use them for different purposes and with different frequency than native-speaking writers (Granger & Rayson, 1998; Hyland, 2002a). Qualitative analysis will show how plural pronouns are used by the Malaysian learners, the focus of Chapter 5. On the other hand, Malaysian learners seem to under-use demonstrative pronouns such as *this*, *these*, *that*, and *those*, which are more common in LOCNESS.

It is also found that there are more adverbs in MCSAW compared to in LOCNESS, especially those expressing place and time (*nowadays*, *sometimes*, *anytime*, and *anywhere*). According to Granger and Rayson (1998: p. 124), learners tend to over-use adverbs that are in relation to place and time, which are considered as 'speech-like adverbs'.

In this section, keywords have been distinguished between lexical and functional ones, and further classified into specific word categories. In turn, the next part involves investigating keywords in terms of their significance and distribution in the corpora.

4.1.2 Significance and distribution

In order to further investigate words that are most common in Malaysian learners' writing, the top 50 keywords are described in more detail in terms of their frequencies and keyness values, indicating how outstanding or statistically salient their frequencies of occurrence are. Table 4.3 presents the top 50 keywords and their respective keyness values. As mentioned in Scott and Tribble (2006: pp. 55-56), "keyness is a quality words may have in a given text or set of texts, suggesting that they are important, they reflect what the text is really about, avoiding trivia and insignificant detail". Those near the top, i.e. those with high keyness values indicate what is statistically more significant in MCSAW (e.g. *Facebook*, *can*, *we* etc). In other words, the higher the keyness, the more statistically significant an item is. As can be seen in Table 4.3, *can* has the largest keyness value after *Facebook* (3773.52) followed by *we* (2727.58). While *Facebook* is expected to be highly 'key' (12517.86) due to being topic-related, the modal verb *can* and pronoun *we* are also found to be significant relative to their usage in the reference corpus. In fact, six of all pronouns are listed in the top 50 keywords of the learner corpus with keyness values ranging from 1364.95 (*our*) to 693.06 (*you*). Together, these results provide empirical support for the decision to focus on modality and pronoun usage in MCSAW.

Table 4.3: Top 50 keywords and their keyness value

N	Key word	Keyness	N	Key word	Keyness
1	facebook	12517.86	26	network	573.66
2	can	3773.52	27	students	512.07
3	we	2727.58	28	online	508.62
4	friends	2155.75	29	nowadays	460.66
5	advantages	1796.04	30	profile	457.46
6	disadvantages	1654.76	31	communicate	450.56
7	our	1364.95	32	with	448.51
8	us	1315.06	33	account	445.95
9	information	1135.40	34	internet	434.18
10	using	1125.44	35	business	429.91
11	use	1117.51	36	medium	421.11
12	social	1095.66	37	hostel	409.62
13	your	1078.05	38	besides	373.98
14	share	894.06	39	communication	373.71
15	people	788.66	40	user	366.13
16	their	761.22	41	news	366.05
17	also	754.34	42	chat	345.28
18	know	707.87	43	student	336.72
19	networking	705.04	44	group	325.76

20	you	693.06	45	stay	314.05
21	time	689.88	46	world	302.81
22	connect	677.99	47	fake	292.95
23	users	655.77	48	easily	279.25
24	get	624.99	49	advantage	275.99
25	friend	604.01	50	conclusion	275.81

However, it should be noted that a word's keyness does not necessarily mean it is distributed evenly throughout the corpus as it may only occur very frequently in one or two texts (Cheng, 2012). To overcome this problem, it is a good idea to examine the distribution, i.e. range of the words distributed in texts (see Chapter 3). Table 4.4 presents the top 50 keywords in terms of their range across the texts of the Malaysian corpus. Out of all 509 texts in MCSAW, the modal verb *can* is found to occur in 97% of them (495 texts). The pronoun *we* is found in 84% (429 texts) of the whole texts followed by words that are deemed topic-related such as *friends* (81%), *advantages* (85%) and *disadvantages* (79%). This in turn, highlights keywords that are not only statistically significant, but they occur in more than one text. Hence, analyses of both statistical significance (keyness) and range suggest that modal verb *can* and personal pronoun *we* are interesting to analyse in more detail.

Table 4.4: Top 50 keywords in terms of their range

N	Key word	Texts	N	Key word	Texts
1	facebook	477	26	network	188
2	can	495	27	students	233
3	we	429	28	online	176
4	friends	414	29	nowadays	256
5	advantages	441	30	profile	116
6	disadvantages	409	31	communicate	175
7	our	370	32	with	474
8	us	347	33	account	170
9	information	343	34	internet	179
10	using	318	35	business	178
11	use	401	36	medium	175
12	social	380	37	hostel	29
13	your	218	38	besides	198
14	share	293	39	communication	177
15	people	454	40	user	118
16	their	459	41	news	172
17	also	457	42	chat	144
18	know	349	43	student	121

19	networking	237	44	group	171
20	you	230	45	stay	126
21	time	375	46	world	339
22	connect	248	47	fake	85
23	users	205	48	easily	196
24	get	333	49	advantage	133
25	friend	199	50	conclusion	238

In addition to investigating keyness value and range of MCSAW keywords, it is also worth examining 'key keyness'. As described in Chapter 3, key keywords indicate keywords which are most frequent over a number of files and therefore, increase the quality of them being text-dependent. Table A4.2 in Appendix shows 25 of 55 items that are considered key keywords.[46] Analysis of key keywords in MCSAW (as shown in Table A4.2) confirms that *can* and *we* are also key keywords. More precisely, *can* is key in 59 texts, i.e. 70% of the 84 key word files, and *we* is key in 43 texts (51%). Other pronouns are also key keywords whereas the remainder of the key keywords are clearly essay or topic-related. The analysis of key keywords adds to the argument that both modality and pronouns are further investigated apart from their significance as well as distribution in the corpus. Section 4.2 thus continues with the focus on modality, while pronouns are examined in Chapter 5.

4.2 Modality

According to Coates (1990: p. 54), "[m]odality has to do with notions such as possibility, necessity, ability, volition, obligation". She further states that in English, "the chief exponents of modality are the modal auxiliaries: *can*, *could*, *will*, *would*, *shall*, *should*, *may*, *might*, *must*, *ought*, *need*, *dare*, and other lexical items to do with possibility, necessity, volition, etc., such as *perhaps*, *possible*, *allow*, *able*, *willing*" (ibid). A modal verb as stated by Mindt (1995: p. 43), "introduces an additional meaning component into the verb phrase", and "generally results in a specification of what is expressed by the main verb". Modal verbs have often been investigated in learner writing from different L1 backgrounds (Gabrielatos & McEnery, 2005; Lee & Chen, 2009; Neff et al., 2003).

[46] This is a result of making keywords lists for every wordlist generated from the total texts in MCSAW, known as a 'database'. The database is "batch-processed to provide key word files" (Scott, 1997: p. 237). In this study, the database resulted in 84 files.

In relation to Malaysian learners, two recent corpus studies (Mohamed Ismail et al., 2013; Mukundan & Khojasteh, 2011) explored and compared the distribution, meanings and contexts of modal use in different types of texts. Both studies found that there were differences observed between authentic English used in natural communicative situations and the kind of English taught in the classroom. Mohamed Ismail et al. (2013: p. 153) investigated modals in a Malaysian corpus of argumentative texts and state that learners used *can* and *will* more frequently than other modal auxiliaries. Mukundan and Khojasteh (2011) compared the distribution of modal auxiliaries in a Malaysian English textbook corpus against the British National Corpus (BNC) and found that the modals were distributed unevenly in the textbooks. According to Mindt (1995), the distribution of modal verbs varies according to text type. For instance, Biber (2006) and Römer (2004) both found that the expression of stance (modal verbs specifically) in written English language textbooks differs considerably from spoken registers. Hence, despite the commonality and importance of modality in English, the discrepancies in which modality are treated in written and spoken discourse can be seen as problematic for learners.

Other difficulties pertaining to the modal system of standard formal English are the similarity in meanings of some modal verbs and the possibility of the same modals being used to express different functions. According to Hyland and Milton (1997: p. 185), "[m]odal expressions are complex for novice writers because they are polypragmatic, that is, they can simultaneously convey a range of different meanings". Kennedy (2002: p. 74) adds that

> modal meanings can be expressed in a number of different ways involving other grammatical and lexical means apart from modal verbs. For example, *You can go outside* and *You have permission to go outside* provide alternative ways of giving permission, but only the former makes use of a modal verb.

In addition, not only do modal verbs have different meanings, their functions can vary according to different contexts, and therefore creating ambiguity (Coates, 1983; Palmer, 1990). In some cases, it can be found that modal expressions in English have also experienced a sense of grammaticalisation, in which modal verbs can be seen to converge with a local variety equivalent (e.g. Chinese, Malay or Tamil) and thus, contribute to the complications that learners have to face (Bao, 2010). In his study of the modal *must*, Bao (2010) discovered that *must* has undergone a change in Singaporean English "in response to pressures from similar modal expressions in the local languages, mainly Chinese and Malay" (p. 1736). His study realised that while *must* has both deontic and epistemic functions in Singaporean English, it is predominantly used in the deontic sense. It is also possible that the Malaysian English local variety has influenced Malaysian learner English as Bao (2010: p. 1736) has shown.

Modals have been described in the literature as expressing degrees of likelihood (epistemic sense) or degrees of obligation, necessity, permission, and volition (deontic sense). In a general sense, modality is related to the speaker's opinion or attitude towards a particular proposition described (Aijmer, 2002). Thus, the use of a modal verb implies that of a speaker/writer's judgement or opinion. In investigating meanings of nine major modal auxiliaries in British English, Coates (1983) firstly discovered that they are used differently in spoken and written registers. She found that for instances like *will* and *would*, the former is used more frequently in speech while the latter occurs more in writing. Interestingly, the same is found with *can* and *could*. While *shall* can be argued to happen more in spoken UK English, the remaining modals (*should*, *must*, *may*, and *might*) are found to be more prevalent in written UK English.

Following Coates (1983), Table 4.5 presents estimate proportions of the nine modals occurring in all corpora, including the ones indicated in Coates (i.e. spoken LLC corpus, written LOB, MCSAW and LOCNESS). Overall, it can be seen that, proportion-wise, the native speakers of LOCNESS use modal verbs relatively similar to the native speakers in LOB, with exceptions for the decrease in the modals *must*, *may*, *might* and *shall* in LOCNESS. In both LOCNESS and LOB, the modals *would*, *will*, and *can* are similarly ranked in descending order, whereas in MCSAW, the most important modals are *can*, *will* and *should*. This reflects the fact that the most frequent modal in MCSAW is *can* with 4,178 occurrences, whereas *would* is highest with 1,461 occurrences in LOCNESS. The percentage of the modal *would* is almost identical in both reference language varieties (21% in LOB and 24% in LOCNESS), but is strikingly low in the Malaysian learner corpus (with about 2%). *Can*, however, is the most important modal in MCSAW with 66%, while it does not constitute more than 20% in either the spoken or written reference language variety. *Will* occurs almost with the same percentage in MCSAW as in written LOB and LOCNESS (19%), but it must be kept in mind that Table 4.5 shows similarities and differences in terms of proportions rather than relative frequencies. In other words, even though the percentage for *will* is similar in MCSAW and LOCNESS, keyness analysis shows that it is over-used in MCSAW (see Table A4.1).

In contrast, Malaysian learners under-use quite a substantial number of modals in their writing (with *would, could, should, must, may, might* and *shall* ranging from 0.1% to 4.4% percentage-wise). The proportion of the modals *might* and *shall* in MCSAW and LOCNESS is low in both corpora, that is, not more than 1.5% and 0.2% respectively. This may be due to the decline of these modals in both corpora compared to the written UK English LOB corpus, which contains informative essays compiled before the 1980s. *Shall* has been found to occur less than other modals in past studies; not occurring more than 1.5 per thousand words in the BNC (Kennedy, 2002: p. 77) and 2.4 per thousand words in the written UK English (LOB) in Coates (1983). It is worth noting that the number gradually declines in both studies (Coates, 1983; Kennedy, 2002), but the opposite is happening to the modals *can* and *will* (Kennedy, 2002: p. 86). This is also true in both MCSAW and LOCNESS, where *will* (1,188) is ranked second most frequent after *can* in MCSAW, and *will* and *can* (both 1,116) are ranked second and third most frequent in LOCNESS.

In sum, Table 4.5 shows that Malaysian learners particularly over-use the modal verb *can* at the expense of other modals distributed in the three reference language varieties. This is in line with past studies, as mentioned earlier (Mohamed Ismail et al., 2013; Mukundan & Khojasteh, 2011), which have identified *can* as being highly frequent in Malaysian learner English writing. However, the huge contrast between the use of *can* and other modals indicate that Malaysian learners tend to over-use the modal verb and may even use it wrongly. One possible reason is the influence of *can* in speech as the modal verb has been shown to occur more frequently in spoken discourse (Mindt, 1995; Coates, 1983).

Table 4.5: Estimate proportions of modals[47]

	Spoken UK English LLC	**Written UK English LOB**	**MCSAW**	**LOCNESS**
Will	24.2	19.3	18.8	18.6
Would	19.9	20.6	1.9	24.3
Can	19.9	14.7	66.1	18.6
Could	11.3	12.0	2.2	10.6
Should	6.3	8.8	4.4	12.8
Must	6.5	7.8	2.4	5.4
May	5.0	9.1	2.6	8.1
Might	4.1	5.3	1.5	1.4
Shall	2.8	2.4	0.1	0.2
Total (%)	100	100	100	100

[47] Also, see Table A4.3 in Appendix for results of the significance test (i.e. keyness) of modal verbs in MCSAW and LOCNESS.

4.2.1 Range of *can* across all L1 backgrounds

As previously discussed, the distribution of keywords presented in Table 4.4 reveals that *can* occurs in 495 texts of the whole 509 texts in MCSAW. This is pertinent in showing that *can* is found to be consistent in majority of MCSAW texts. A further analysis of the detailed consistency or range of the modal verb *can* in both learner and reference language variety corpora enables more comparisons to be made, especially with regard to stylistic reasons, revealing that *can* occurs across 97% of the Malaysian corpus compared to only 16% in the reference corpus. Unlike native speakers in LOCNESS, this shows that *can* appears to be the preferred marker of modality for these Malaysian learners.

Further analysis of *can* is also carried out between the three separate learner groups which constitute MCSAW, notably the Malay, Chinese and Indian learners. This in turn, demonstrates the distribution of *can* in each separate learner group writing respectively. Figure 4.1 shows the portions of a pie chart that graphically represent *can* used among Malay, Chinese, and Indian learners of MCSAW. It can be seen that majority of *can* usage (86%) is by the Malay learners, followed by the Chinese learners (12%) and to a lesser extent: 2% by the Indian learners. When compared to the overall distribution of texts in MCSAW according to the different L1 groups as shown in Figure 4.2, the results are relatively proportional. This signals that regardless of L1 background, it may be that all learners in MCSAW have similar problems with the overuse of *can*. Therefore, contrary to past research (Gabrielatos & McEnery, 2005), these findings suggest that the use of modal verb *can* in Malaysian learner English writing may point to two possibilities: either this occurrence is not purely indicative of learners' L1 influence (but rather influenced by another factor altogether) or they are all equally influenced by their L1. Further, as noted previously, differences in language use may be associated with competency rather than a writer's L1 background. It is also important to note that although the Malay group of learners appear to use slightly more *can* in their texts (4% difference), it is too small a difference to make any significant claims. Further research could however, be undertaken to resolve this issue by collecting more evenly texts by the three major learner groups in Malaysia.

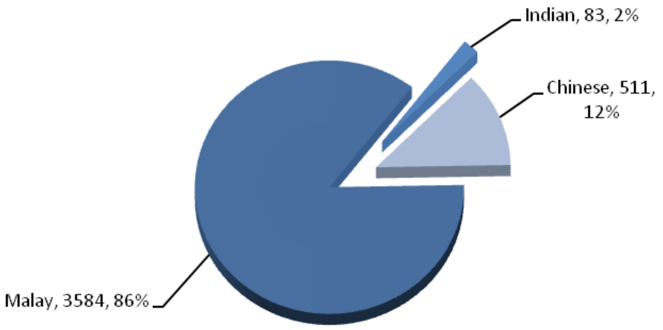


Figure 4.1: *Can* occurrences according to L1 groups in MCSAW

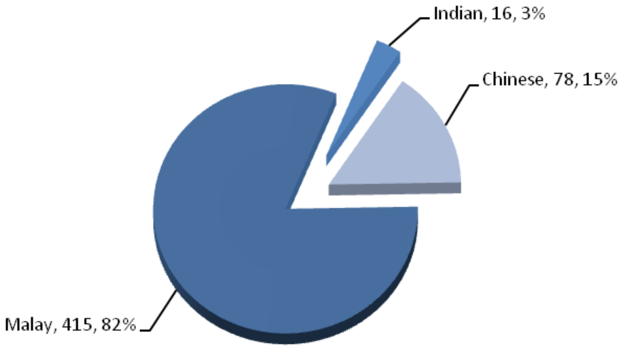
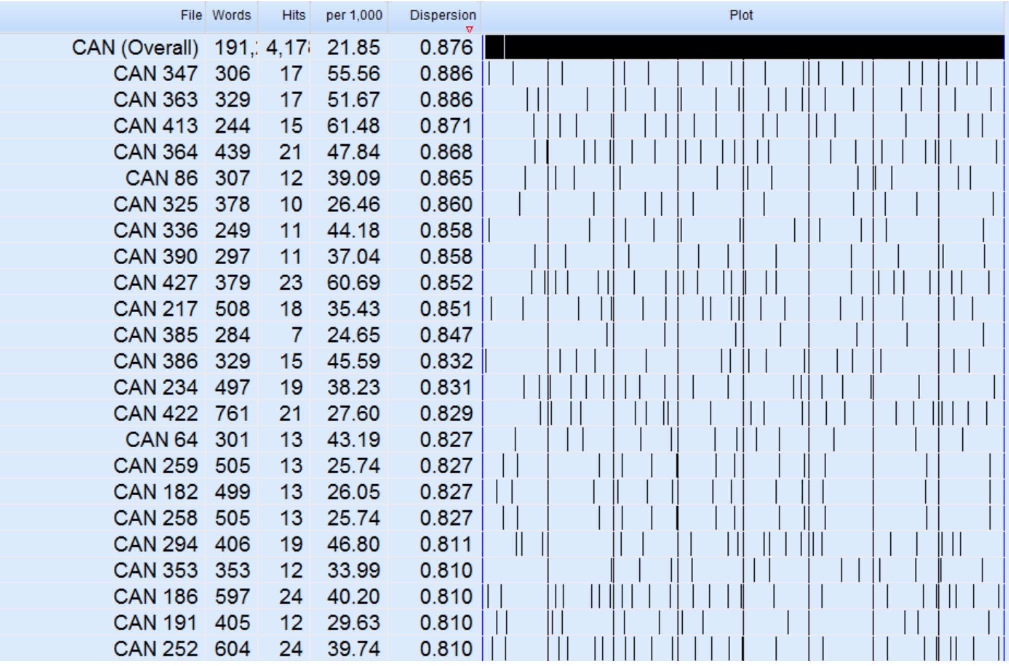


Figure 4.2: Distribution of texts according to L1 groups in MCSAW

4.2.2 Dispersion plot

Another important investigation is to examine the plot of the modal *can* in texts. This is insightful because it allows for *can* to be searched in the corpus to see where mention is made most in each text. In addition, it promotes the noticing of linguistic patterning that could be representative of a particular genre structure. Figure 4.3 presents a sample of the plot diagram that illustrates the scattering of *can* in a number of Malaysian learner English texts. The plot shows a dispersion value, in which the statistics give mathematical support to indicate whether *can* is evenly distributed. It ranges from 0 to 1, with 0.9 or 1 suggesting very uniform dispersion and 0 or 0.1 suggesting irregular distribution (Scott, 2015). An examination of each of the texts where *can* occurs in MCSAW indicates that only 22 out of the 496 texts in which *can* occurs had a dispersion value close to 0.1. The remainder of texts on the other hand, showed a dispersion value above 0.1, and 368 texts specifically were above 0.5. In addition, the overall dispersion value for *can* in the 496 texts it occurs in is 0.876, which is close to 0.9, resulting in a uniform plot as shown in Figure 4.3.

Can is seen to occur in almost all parts of the essays, regardless of position, with only fewer occurrences towards the end. Following Hyland's (1990) description of the stages of a typical argumentative essay (as discussed in Chapter 3), the dispersion of *can* in MCSAW in turn means that learners use *can* in all parts of their essays including the thesis, argument and conclusion. Similarly, this may indicate that the use of *can* is widespread in describing the discourse functions pertinent to each of the essay parts mentioned by Hyland (1990: p. 69), namely introducing the proposition of argument, discussing the argument, and synthesising the discussion as well as affirming the validity of the proposition. While it is not the focus of this chapter (or this thesis) to examine *can* in terms of a genre analysis, the employment of the plot function in WordSmith Tools illustrates the overuse of *can* as extensive throughout texts in the corpus. In examining the use of *can* further, collocational analysis is discussed next.



File	Words	Hits	per 1,000	Dispersion	Plot
CAN (Overall)	191,:	4,17:	21.85	0.876	
CAN 347	306	17	55.56	0.886	
CAN 363	329	17	51.67	0.886	
CAN 413	244	15	61.48	0.871	
CAN 364	439	21	47.84	0.868	
CAN 86	307	12	39.09	0.865	
CAN 325	378	10	26.46	0.860	
CAN 336	249	11	44.18	0.858	
CAN 390	297	11	37.04	0.858	
CAN 427	379	23	60.69	0.852	
CAN 217	508	18	35.43	0.851	
CAN 385	284	7	24.65	0.847	
CAN 386	329	15	45.59	0.832	
CAN 234	497	19	38.23	0.831	
CAN 422	761	21	27.60	0.829	
CAN 64	301	13	43.19	0.827	
CAN 259	505	13	25.74	0.827	
CAN 182	499	13	26.05	0.827	
CAN 258	505	13	25.74	0.827	
CAN 294	406	19	46.80	0.811	
CAN 353	353	12	33.99	0.810	
CAN 186	597	24	40.20	0.810	
CAN 191	405	12	29.63	0.810	
CAN 252	604	24	39.74	0.810	

Figure 4.3: Dispersion plot for *can* occurrences in MCSAW

4.2.3 Collocation comparison

As discussed in Chapter 3, collocation shows the co-occurrence of two words with a frequency above chance, which is statistically calculated. More specifically, collocational analysis indicates "[p]atterns of association – how lexical items tend to co-occur – [that] are built up over large amounts of text and are often unavailable to intuition or conscious awareness" (Hunston, 2002a: p. 109). Exploring collocates of a word thus, reveals the common lexical and grammatical patterns of a co-occurrence. Table 4.6 and Table 4.7 present collocates for *can* in both the Malaysian learner corpus and reference language variety using t-score and MI respectively, with settings for both t-score and MI set at a 5:5 span. As mentioned in Chapter 3, it is important to make comparisons between the two measures as Hunston (2002b: p. 73) notes, they show different things: "MI-score is a measure of strength of collocation, [whereas] t-score is a measure of certainty of collocation". This accounts for the more grammatical words in Table 4.6 (e.g. *also*, *we*, *that*, *with*), and more lexical words in Table 4.7 (e.g. *leverage*, *concluded*, *threatening*, *duty*). Also, there is a tendency for frequent words to be collocates with the highest t-scores, while collocates with the highest MI-scores tends to be less frequent words with restricted collocation (Gabrielatos & Baker, 2006). Based on these statistical significance tests, some observations can be made with regard to collocates listed by both statistical measures, starting with some discussion on *can* collocates using t-score followed by the MI-score and shared collocates in both corpora.

Table 4.6: Collocates for *can* in both MCSAW and LOCNESS using t-score

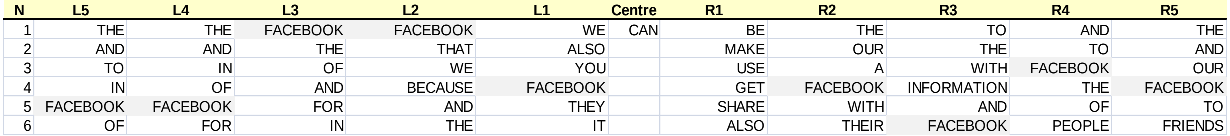
Only in MCSAW	**Only in LOCNESS**	**In both**
also (16.45), we (15.11), that (14.75), with (14.58), be (14.4), from (13.33), our (13.3), get (13.27), and (13.24), people (13.2), share (13.05), a (13.04), facebook (12.68), use (12.6), friends (12.58), information (12.38), connect (12.36), in (12.07), to (12.05), of (12.03), make (11.95), it (11.84), for (11.83), the (11.57), they (11.4), their (11.37), as (11.36), you (10.85), because (10.52), about (10.35), know (10.19), so (10.11), this (10.11), help (10.1), using (10.08), or (10.03), many (10.02), find (10.01), by (9.97), them (9.65), besides (9.54), other (9.38), easily (9.33), us (9.31), is (9.23), your (9.08), group (9.05), more (8.77)	lead (3.2), seen (3.1), said (2.67), so (2.61), only (2.6), any (2.57), see (2.48), we (2.48), do (2.46), same (2.4), from (2.39), how (2.34), make (2.34), done (2.32), through (2.31), very (2.31), become (2.3), way (2.3), take (2.29), cause (2.27), what (2.27), no (2.27), human (2.25), find (2.25), come (2.25), disease (2.25), often (2.24), you (2.21), some (2.2), much (2.2), afford (2.18), produce (2.17), happen (2.14), start (2.13), understand (2.13), then (2.09), where (2.07), than (2.07), situation (2.06), up (2.06), therefore (2.06), if (2.06), like (2.04), death (2.03), though (2.03), now (2.01), which (2.01)	we, from, make, you, so, find

Table 4.7: Collocates for *can* in both MCSAW and LOCNESS using MI score

Only in MCSAW	**Only in LOCNESS**	**In both**
concluded (5.56), threatening (5.56), what's (5.56), leverage (5.56), duty (5.3), avoided (5.3), walk (5.21), feedback (5.16), conclude (5.12), publish (5.1), stalk (5.06), experiences (5.05), unknown (5.02), later (5.01), obtained (4.98), independent (4.98), brief (4.91), engine (4.88), download (4.88), track (4.87), profit (4.82), burdening (4.81), trouble (4.74), blogs (4.73), creative (4.71), article (4.71), obtain (4.71), immediately (4.69), stop (4.69), stressful (4.69), gather (4.68), interests (4.68), videos (4.66), solve (4.65), enjoy (4.64), exchange (4.61), hope (4.61), save (4.59), directly (4.56), maintain (4.56), anybody (4.56), learn (4.56), found (4.56), freely (4.48), seeing (4.47), unlimited (4.47), third (4.46), different (4.45), bond (4.45), power (4.45)	sympathise (6.48), afford (5.33), produce (5.04), damage (4.93), contract (4.89), possibly (4.82), enjoy (4.8), hold (4.76), lead (4.7), sometimes (4.69), compete (4.51), genes (4.36), travel (4.31), benefit (4.24), done (4.2), easily (4.13), deal (4.09), tell (4.09), start (4.06), understand (4.06), sure (4.04), improve (4.04), decisions (4.02), later (3.99), seen (3.96), said (3.96), happiness (3.93), works (3.89), humans (3.89), influence (3.82), cause (3.8), situation (3.7), anything (3.69), program (3.65), either (3.65), disease (3.61), found (3.6), nothing (3.56), information (3.53), shown (3.51), though (3.42), prove (3.38), find (3.37), come (3.36), effect (3.34), often (3.32), buy (3.31), patient (3.26), control (3.26), suffering (3.24)	later, enjoy, found

4.2.3.1 Collocates identified using t-score

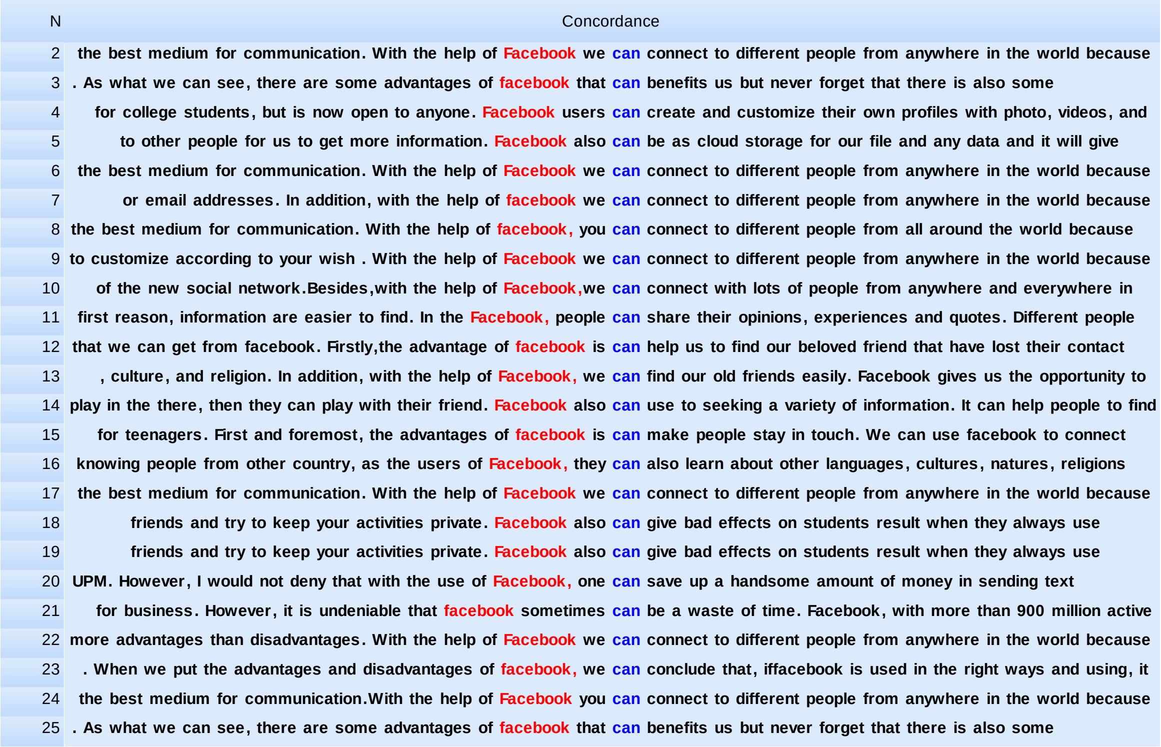
Firstly, it can be seen from Table 4.6 that there are a number of collocates referring to the topics in MCSAW, i.e. *Facebook*, *information*, *connect*. These collocates indicate the co-occurrence of the modal verb *can* in connection with the content of texts in MCSAW – especially Facebook. Further inspection reveals that for *Facebook*, the collocate is seen to frequently co-occur with *can* in the following positions, as is shown in Figure 4.4:



N	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	THE	THE	FACEBOOK	FACEBOOK	WE	CAN	BE	THE	TO	AND	THE
2	AND	AND	THE	THAT	ALSO		MAKE	OUR	THE	TO	AND
3	TO	IN	OF	WE	YOU		USE	A	WITH	FACEBOOK	OUR
4	IN	OF	AND	BECAUSE	FACEBOOK		GET	FACEBOOK	INFORMATION	THE	FACEBOOK
5	FACEBOOK	FACEBOOK	FOR	AND	THEY		SHARE	WITH	AND	OF	TO
6	OF	FOR	IN	THE	IT		ALSO	THEIR	FACEBOOK	PEOPLE	FRIENDS

Figure 4.4: Collocate positions for *Facebook*

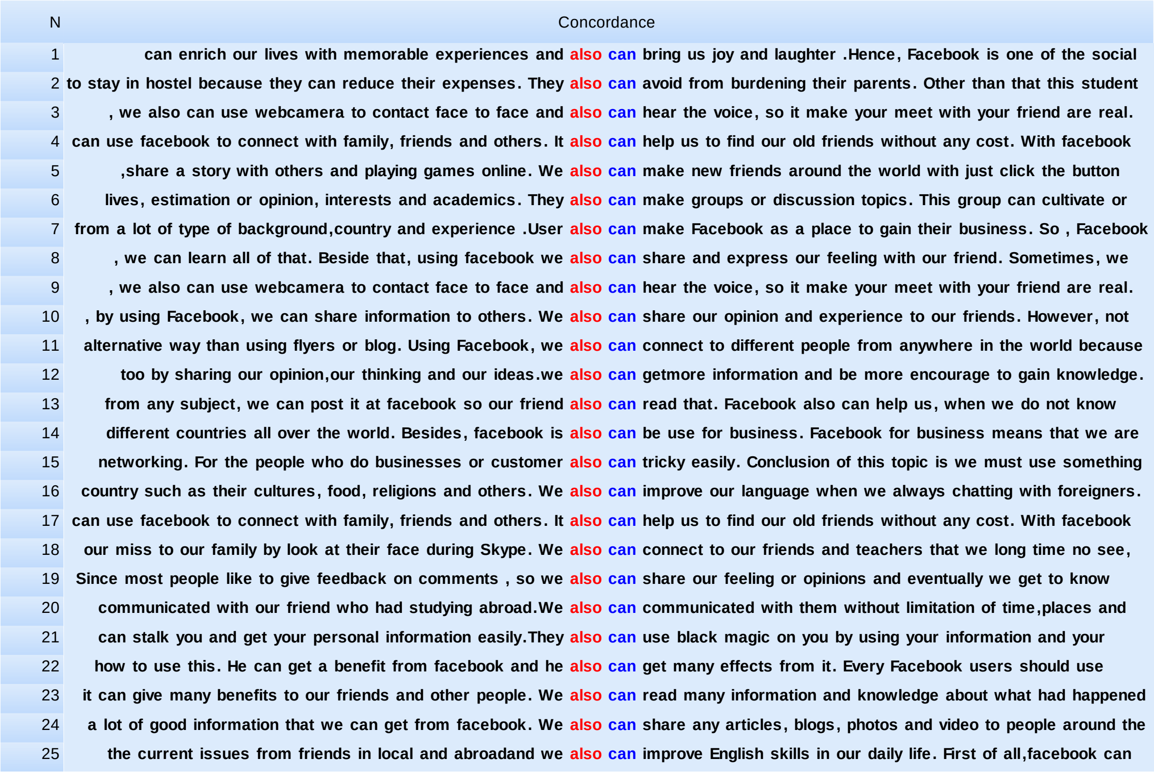
Among all the positions in which *Facebook* co-occurs with *can*, it is found that the collocate *Facebook* is more frequent in the second left (henceforth, L2) position (359 times). Figure 4.5 presents several concordance lines for this pattern. It can be seen that some of the patterns include the use of Facebook in adverbial phrases (e.g. *with the help of Facebook*, *in the Facebook, as the users of Facebook*), indicating that the modal verb *can* follows these phrases in relation to the context of Facebook. It is also found that several lines show that the phrase *connect to different people from anywhere in/all around the world* is used repeatedly. Closer inspection however, reveals that the lines are taken from different texts. This could be indicative of a commonly taught phrase in the classroom, which is likely to suggest a form of scaffolding in the writing processes. Alternatively, it could indicate copying by the students. Other patterns for the use of *Facebook* as being a certain collocate of *can* indicate *Facebook* as a frequent subject in the texts. Examples include *Facebook users can* and *Facebook also can*.



N	Concordance
2	the best medium for communication. With the help of Facebook we can connect to different people from anywhere in the world because
3	. As what we can see, there are some advantages of facebook that can benefits us but never forget that there is also some
4	for college students, but is now open to anyone. Facebook users can create and customize their own profiles with photo, videos, and
5	to other people for us to get more information. Facebook also can be as cloud storage for our file and any data and it will give
6	the best medium for communication. With the help of Facebook we can connect to different people from anywhere in the world because
7	or email addresses. In addition, with the help of facebook we can connect to different people from anywhere in the world because
8	the best medium for communication. With the help of facebook, you can connect to different people from all around the world because
9	to customize according to your wish . With the help of Facebook we can connect to different people from anywhere in the world because
10	of the new social network.Besides,with the help of Facebook,we can connect with lots of people from anywhere and everywhere in
11	first reason, information are easier to find. In the Facebook, people can share their opinions, experiences and quotes. Different people
12	that we can get from facebook. Firstly,the advantage of facebook is can help us to find our beloved friend that have lost their contact
13	, culture, and religion. In addition, with the help of Facebook, we can find our old friends easily. Facebook gives us the opportunity to
14	play in the there, then they can play with their friend. Facebook also can use to seeking a variety of information. It can help people to find
15	for teenagers. First and foremost, the advantages of facebook is can make people stay in touch. We can use facebook to connect
16	knowing people from other country, as the users of Facebook, they can also learn about other languages, cultures, natures, religions
17	the best medium for communication. With the help of Facebook we can connect to different people from anywhere in the world because
18	friends and try to keep your activities private. Facebook also can give bad effects on students result when they always use
19	friends and try to keep your activities private. Facebook also can give bad effects on students result when they always use
20	UPM. However, I would not deny that with the use of Facebook, one can save up a handsome amount of money in sending text
21	for business. However, it is undeniable that facebook sometimes can be a waste of time. Facebook, with more than 900 million active
22	more advantages than disadvantages. With the help of Facebook we can connect to different people from anywhere in the world because
23	. When we put the advantages and disadvantages of facebook, we can conclude that, iffacebook is used in the right ways and using, it
24	the best medium for communication.With the help of Facebook you can connect to different people from anywhere in the world because
25	. As what we can see, there are some advantages of facebook that can benefits us but never forget that there is also some

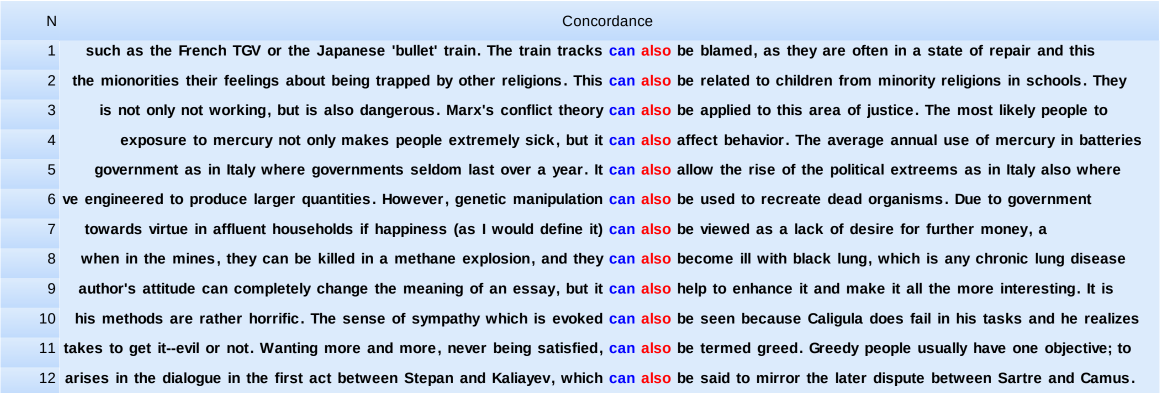
Figure 4.5: Concordance lines for *Facebook* co-occurring with *can* in L2 position

The most frequent collocates in terms of t-score presented in Table 4.6 are *also* (16.45), *we* (15.11), *that* (14.75), *with* (14.58), and *be* (14.4). It is found that *also* co-occurs 417 times to the immediate left of *can*, resulting in the *also can* cluster as shown in Figure 4.6. The phrase *also can* seems to occur following the conjunction *and*, the pronouns *they*, *it*, *we*, *he*, and the nouns *user*, *friend*, and *customer* – implying additional information to the previous sentence, which instead, could be replaced with conjunctions such as *in addition*, *moreover*, *furthermore* etc. One possible explanation lies in the spoken feature of *also can* in colloquial Malaysian English local variety (or Manglish). However, *also can* appears only once in the reference language variety, where the use of *can also* is found to be more prevalent, as shown in Figure 4.7. It is also noteworthy to add that there is a difference in the use of the passive structure in Figure 4.7 (e.g. *can also be blamed*/*be related*, *be applied* etc.) compared to the active use in Figure 4.6 (e.g. *also can bring*/*avoid*/*hear* etc).



N	Concordance
1	can enrich our lives with memorable experiences and also can bring us joy and laughter .Hence, Facebook is one of the social
2	to stay in hostel because they can reduce their expenses. They also can avoid from burdening their parents. Other than that this student
3	, we also can use webcamera to contact face to face and also can hear the voice, so it make your meet with your friend are real.
4	can use facebook to connect with family, friends and others. It also can help us to find our old friends without any cost. With facebook
5	,share a story with others and playing games online. We also can make new friends around the world with just click the button
6	lives, estimation or opinion, interests and academics. They also can make groups or discussion topics. This group can cultivate or
7	from a lot of type of background,country and experience .User also can make Facebook as a place to gain their business. So , Facebook
8	, we can learn all of that. Beside that, using facebook we also can share and express our feeling with our friend. Sometimes, we
9	, we also can use webcamera to contact face to face and also can hear the voice, so it make your meet with your friend are real.
10	, by using Facebook, we can share information to others. We also can share our opinion and experience to our friends. However, not
11	alternative way than using flyers or blog. Using Facebook, we also can connect to different people from anywhere in the world because
12	too by sharing our opinion,our thinking and our ideas.we also can getmore information and be more encourage to gain knowledge.
13	from any subject, we can post it at facebook so our friend also can read that. Facebook also can help us, when we do not know
14	different countries all over the world. Besides, facebook is also can be use for business. Facebook for business means that we are
15	networking. For the people who do businesses or customer also can tricky easily. Conclusion of this topic is we must use something
16	country such as their cultures, food, religions and others. We also can improve our language when we always chatting with foreigners.
17	can use facebook to connect with family, friends and others. It also can help us to find our old friends without any cost. With facebook
18	our miss to our family by look at their face during Skype. We also can connect to our friends and teachers that we long time no see,
19	Since most people like to give feedback on comments , so we also can share our feeling or opinions and eventually we get to know
20	communicated with our friend who had studying abroad.We also can communicated with them without limitation of time,places and
21	can stalk you and get your personal information easily.They also can use black magic on you by using your information and your
22	how to use this. He can get a benefit from facebook and he also can get many effects from it. Every Facebook users should use
23	it can give many benefits to our friends and other people. We also can read many information and knowledge about what had happened
24	a lot of good information that we can get from facebook. We also can share any articles, blogs, photos and video to people around the
25	the current issues from friends in local and abroadand we also can improve English skills in our daily life. First of all,facebook can

Figure 4.6: Concordance lines for *also* co-occurring with *can*

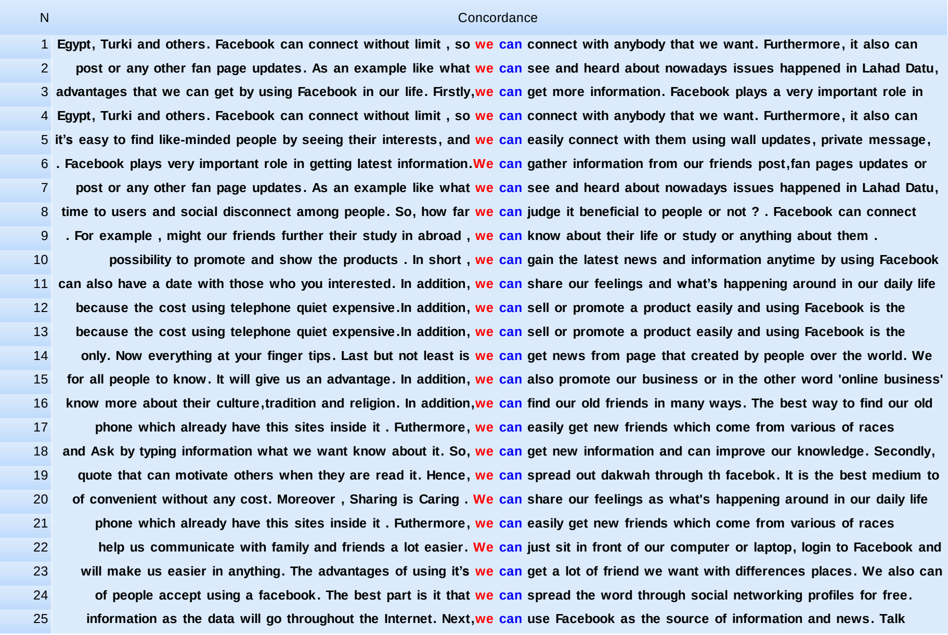


N	Concordance
1	such as the French TGV or the Japanese 'bullet' train. The train tracks can also be blamed, as they are often in a state of repair and this
2	the mionorities their feelings about being trapped by other religions. This can also be related to children from minority religions in schools. They
3	is not only not working, but is also dangerous. Marx's conflict theory can also be applied to this area of justice. The most likely people to
4	exposure to mercury not only makes people extremely sick, but it can also affect behavior. The average annual use of mercury in batteries
5	government as in Italy where governments seldom last over a year. It can also allow the rise of the political extreems as in Italy also where
6	ve engineered to produce larger quantities. However, genetic manipulation can also be used to recreate dead organisms. Due to government
7	towards virtue in affluent households if happiness (as I would define it) can also be viewed as a lack of desire for further money, a
8	when in the mines, they can be killed in a methane explosion, and they can also become ill with black lung, which is any chronic lung disease
9	author's attitude can completely change the meaning of an essay, but it can also help to enhance it and make it all the more interesting. It is
10	his methods are rather horrific. The sense of sympathy which is evoked can also be seen because Caligula does fail in his tasks and he realizes
11	takes to get it--evil or not. Wanting more and more, never being satisfied, can also be termed greed. Greedy people usually have one objective; to
12	arises in the dialogue in the first act between Stepan and Kaliayev, which can also be said to mirror the later dispute between Sartre and Camus.

Figure 4.7: Concordance lines for *can also* in LOCNESS

We is also most frequently found to co-occur immediately preceding *can* (1,249 times) as is shown in Figure 4.8. Constructions *we can* prove to be the most statistically significant among Malaysian learners (1,260 times). This is similar to that which is found in Neff et al. (2003) in all learner texts (Italians, French, and Spanish), except for German writers. Also, the cluster *we can* in MCSAW appears to be connected to a number of lexical verbs, identified by Granger and Paquot (2009) as over-used in most learner texts (ICLE) compared to the native language variety, such as *get* (13.27), *use* (12.6), *make* (11.95), *know* (10.19), *help* (10.1), and *using* (10.08). These lexical verbs that co-occur frequently with the cluster *we can* are claimed to be high frequency words (Granger & Paquot, 2009).

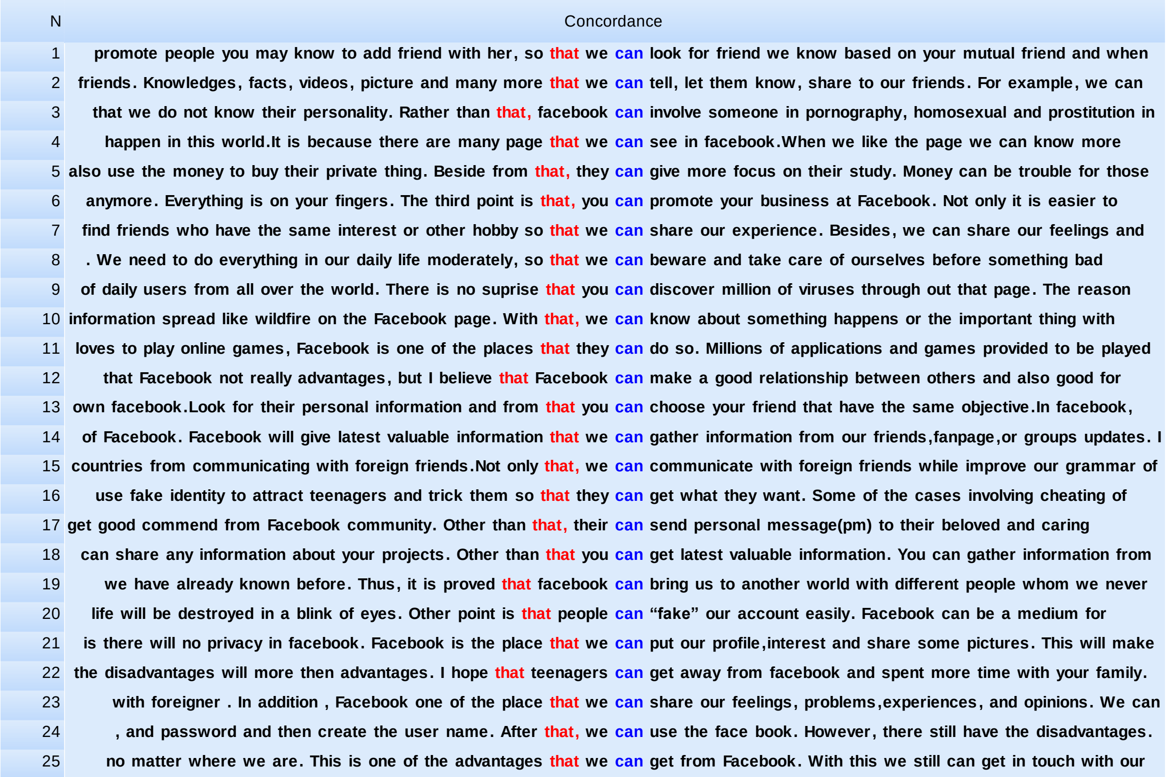
Besides topic-related verbs like *share* (13.05), *find* (10.01), and *connect* (12.36), most of the lexical verbs are marked as typical of conversation and usually uncommon in academic texts (Granger & Paquot, 2009: pp. 202-203). More importantly, it is found that *we* in this pattern, functions mostly as the inclusive *we*. These clusters indicate pragmatic function of including the reader in the writer's discourse community and assuming that the information presented is common knowledge, instead of constructing a more impersonal reader-in-the-text stance, such as 'it might be argued' (Neff et al., 2004: p. 563), which does not oblige the reader to take on board the proposition. This usage pattern of *we can* in relation to argumentative essays is in fact found to be a feature in both MCSAW and LOCNESS writing, since *we* is also a collocate of *can* in LOCNESS.



N	Concordance
1	Egypt, Turki and others. Facebook can connect without limit , so we can connect with anybody that we want. Furthermore, it also can
2	post or any other fan page updates. As an example like what we can see and heard about nowadays issues happened in Lahad Datu,
3	advantages that we can get by using Facebook in our life. Firstly,we can get more information. Facebook plays a very important role in
4	Egypt, Turki and others. Facebook can connect without limit , so we can connect with anybody that we want. Furthermore, it also can
5	it's easy to find like-minded people by seeing their interests, and we can easily connect with them using wall updates, private message,
6	. Facebook plays very important role in getting latest information.We can gather information from our friends post,fan pages updates or
7	post or any other fan page updates. As an example like what we can see and heard about nowadays issues happened in Lahad Datu,
8	time to users and social disconnect among people. So, how far we can judge it beneficial to people or not ? . Facebook can connect
9	. For example , might our friends further their study in abroad , we can know about their life or study or anything about them .
10	possibility to promote and show the products . In short , we can gain the latest news and information anytime by using Facebook
11	can also have a date with those who you interested. In addition, we can share our feelings and what's happening around in our daily life
12	because the cost using telephone quiet expensive.In addition, we can sell or promote a product easily and using Facebook is the
13	because the cost using telephone quiet expensive.In addition, we can sell or promote a product easily and using Facebook is the
14	only. Now everything at your finger tips. Last but not least is we can get news from page that created by people over the world. We
15	for all people to know. It will give us an advantage. In addition, we can also promote our business or in the other word 'online business'
16	know more about their culture,tradition and religion. In addition,we can find our old friends in many ways. The best way to find our old
17	phone which already have this sites inside it . Futhermore, we can easily get new friends which come from various of races
18	and Ask by typing information what we want know about it. So, we can get new information and can improve our knowledge. Secondly,
19	quote that can motivate others when they are read it. Hence, we can spread out dakwah through th facebok. It is the best medium to
20	of convenient without any cost. Moreover , Sharing is Caring . We can share our feelings as what's happening around in our daily life
21	phone which already have this sites inside it . Futhermore, we can easily get new friends which come from various of races
22	help us communicate with family and friends a lot easier. We can just sit in front of our computer or laptop, login to Facebook and
23	will make us easier in anything. The advantages of using it's we can get a lot of friend we want with differences places. We also can
24	of people accept using a facebook. The best part is it that we can spread the word through social networking profiles for free.
25	information as the data will go throughout the Internet. Next,we can use Facebook as the source of information and news. Talk

Figure 4.8: Concordance lines for *we* co-occurring with *can*

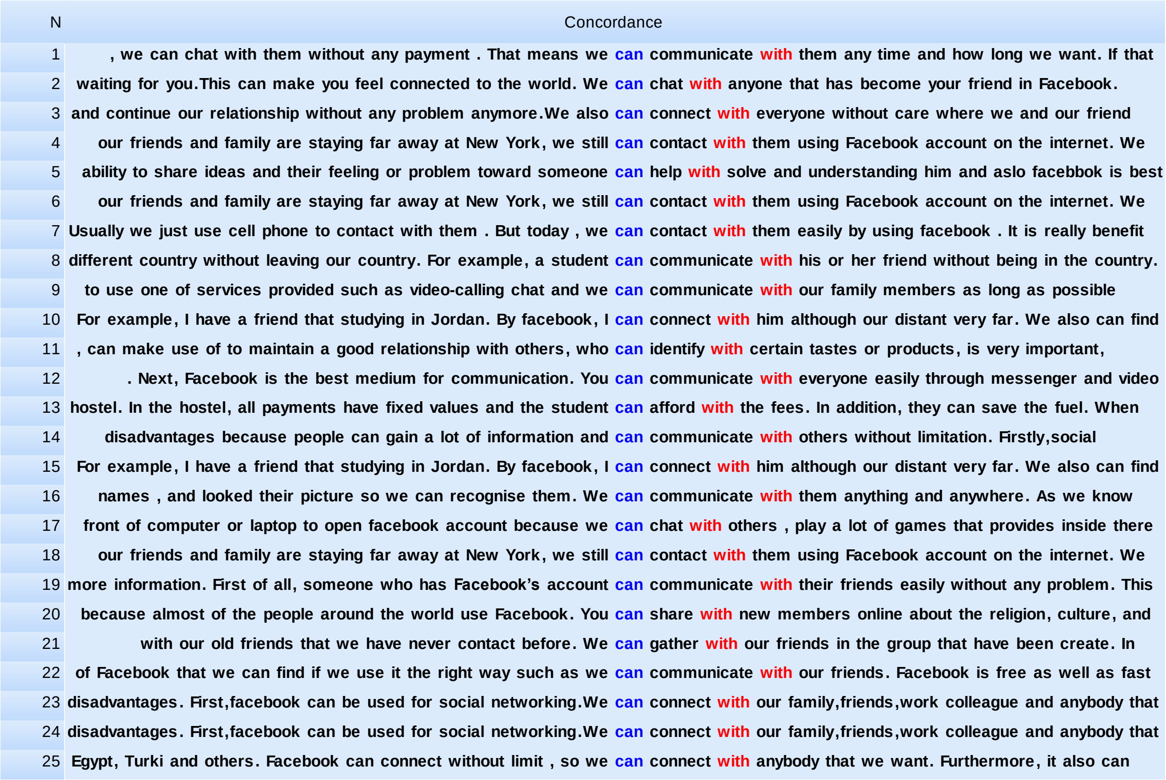
Other frequent collocates of *can* are *that* (14.75), *with* (14.58), and *be* (14.4). It is found that the highly co-occurring *that* in the L2 position[48] of *can* (261 times) indicate a number of -*that* phrases such as *so that*, *with that*, *not only that*, *other than that*, *after that*, *besides that*, and *apart from that* before the use of *can* (as shown in Figure 4.9). The collocate *with* co-occurs most frequently in the R2 position (169 times), mostly conveying a relationship with something/someone that depends on the co-occurring main verb (as in *can communicate/chat/connect* ***with*** *them/anyone/everyone*), whereas the collocate *be* co-occurs most frequently immediately to the right of *can* (356 times), resulting in the *can be* cluster, which signifies most of the passive constructions like *can be received*, *can be seen*, *can be used* etc. (Figure 4.11). It is also important to note that some of the instances are similar to one another (consider concordance lines 23 and 24 in Figure 4.10). Although further investigation reveals that they are from different texts, evidence of similar forms of sentences could suggest a possible prompt learnt in the classroom or the copying of text, as suggested previously.



N	Concordance
1	promote people you may know to add friend with her, so that we can look for friend we know based on your mutual friend and when
2	friends. Knowledges, facts, videos, picture and many more that we can tell, let them know, share to our friends. For example, we can
3	that we do not know their personality. Rather than that, facebook can involve someone in pornography, homosexual and prostitution in
4	happen in this world.It is because there are many page that we can see in facebook.When we like the page we can know more
5	also use the money to buy their private thing. Beside from that, they can give more focus on their study. Money can be trouble for those
6	anymore. Everything is on your fingers. The third point is that, you can promote your business at Facebook. Not only it is easier to
7	find friends who have the same interest or other hobby so that we can share our experience. Besides, we can share our feelings and
8	. We need to do everything in our daily life moderately, so that we can beware and take care of ourselves before something bad
9	of daily users from all over the world. There is no suprise that you can discover million of viruses through out that page. The reason
10	information spread like wildfire on the Facebook page. With that, we can know about something happens or the important thing with
11	loves to play online games, Facebook is one of the places that they can do so. Millions of applications and games provided to be played
12	that Facebook not really advantages, but I believe that Facebook can make a good relationship between others and also good for
13	own facebook.Look for their personal information and from that you can choose your friend that have the same objective.In facebook,
14	of Facebook. Facebook will give latest valuable information that we can gather information from our friends,fanpage,or groups updates. I
15	countries from communicating with foreign friends.Not only that, we can communicate with foreign friends while improve our grammar of
16	use fake identity to attract teenagers and trick them so that they can get what they want. Some of the cases involving cheating of
17	get good commend from Facebook community. Other than that, their can send personal message(pm) to their beloved and caring
18	can share any information about your projects. Other than that you can get latest valuable information. You can gather information from
19	we have already known before. Thus, it is proved that facebook can bring us to another world with different people whom we never
20	life will be destroyed in a blink of eyes. Other point is that people can “fake” our account easily. Facebook can be a medium for
21	is there will no privacy in facebook. Facebook is the place that we can put our profile,interest and share some pictures. This will make
22	the disadvantages will more then advantages. I hope that teenagers can get away from facebook and spent more time with your family.
23	with foreigner . In addition , Facebook one of the place that we can share our feelings, problems,experiences, and opinions. We can
24	, and password and then create the user name. After that, we can use the face book. However, there still have the disadvantages.
25	no matter where we are. This is one of the advantages that we can get from Facebook. With this we still can get in touch with our

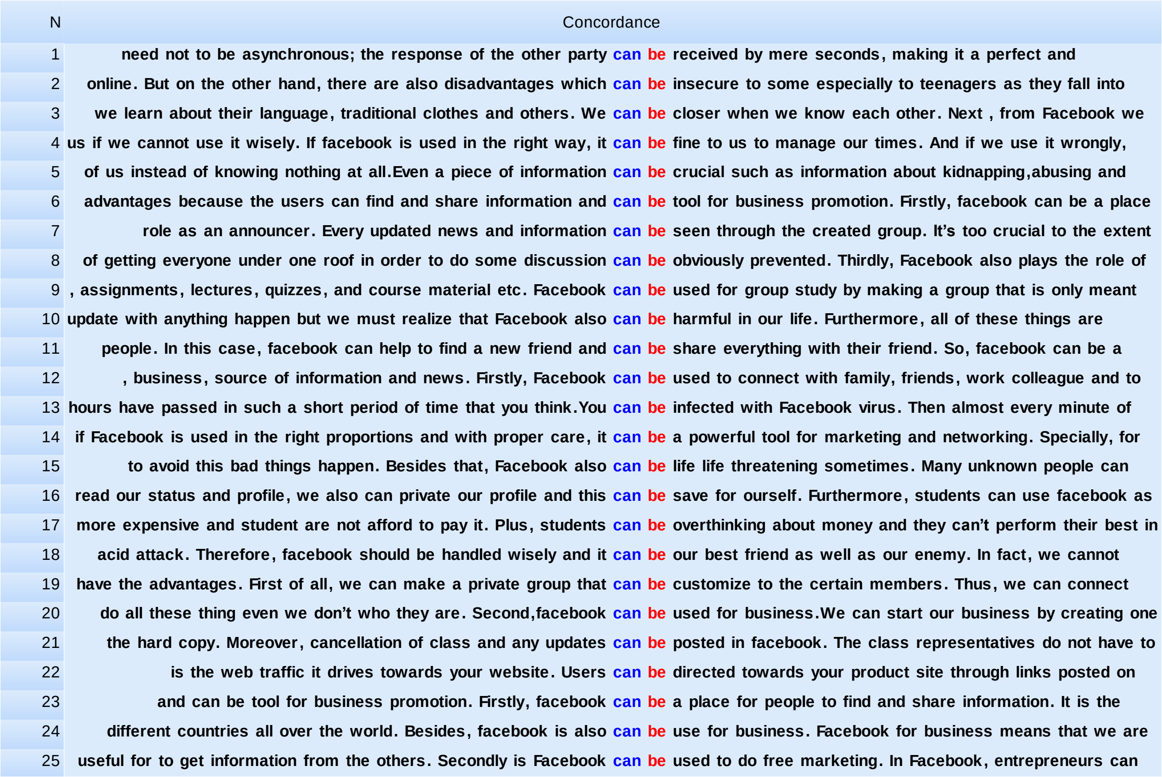
Figure 4.9: Concordance lines for collocate *that*

[48] Collocates occurring immediately to the left of a word will be identified as L1, while R1 signals collocates occurring immediately to the right of the word. Collocates positioned two words to the left of a word therefore is identified as L2 and so on. This will also be used throughout analysing collocates in the next chapter.



N	Concordance
1	, we can chat with them without any payment . That means we can communicate with them any time and how long we want. If that
2	waiting for you.This can make you feel connected to the world. We can chat with anyone that has become your friend in Facebook.
3	and continue our relationship without any problem anymore.We also can connect with everyone without care where we and our friend
4	our friends and family are staying far away at New York, we still can contact with them using Facebook account on the internet. We
5	ability to share ideas and their feeling or problem toward someone can help with solve and understanding him and aslo facebbok is best
6	our friends and family are staying far away at New York, we still can contact with them using Facebook account on the internet. We
7	Usually we just use cell phone to contact with them . But today , we can contact with them easily by using facebook . It is really benefit
8	different country without leaving our country. For example, a student can communicate with his or her friend without being in the country.
9	to use one of services provided such as video-calling chat and we can communicate with our family members as long as possible
10	For example, I have a friend that studying in Jordan. By facebook, I can connect with him although our distant very far. We also can find
11	, can make use of to maintain a good relationship with others, who can identify with certain tastes or products, is very important,
12	. Next, Facebook is the best medium for communication. You can communicate with everyone easily through messenger and video
13	hostel. In the hostel, all payments have fixed values and the student can afford with the fees. In addition, they can save the fuel. When
14	disadvantages because people can gain a lot of information and can communicate with others without limitation. Firstly,social
15	For example, I have a friend that studying in Jordan. By facebook, I can connect with him although our distant very far. We also can find
16	names , and looked their picture so we can recognise them. We can communicate with them anything and anywhere. As we know
17	front of computer or laptop to open facebook account because we can chat with others , play a lot of games that provides inside there
18	our friends and family are staying far away at New York, we still can contact with them using Facebook account on the internet. We
19	more information. First of all, someone who has Facebook's account can communicate with their friends easily without any problem. This
20	because almost of the people around the world use Facebook. You can share with new members online about the religion, culture, and
21	with our old friends that we have never contact before. We can gather with our friends in the group that have been create. In
22	of Facebook that we can find if we use it the right way such as we can communicate with our friends. Facebook is free as well as fast
23	disadvantages. First,facebook can be used for social networking.We can connect with our family,friends,work colleague and anybody that
24	disadvantages. First,facebook can be used for social networking.We can connect with our family,friends,work colleague and anybody that
25	Egypt, Turki and others. Facebook can connect without limit , so we can connect with anybody that we want. Furthermore, it also can

Figure 4.10: Concordance lines for collocate *with*



N	Concordance
1	need not to be asynchronous; the response of the other party can be received by mere seconds, making it a perfect and
2	online. But on the other hand, there are also disadvantages which can be insecure to some especially to teenagers as they fall into
3	we learn about their language, traditional clothes and others. We can be closer when we know each other. Next , from Facebook we
4	us if we cannot use it wisely. If facebook is used in the right way, it can be fine to us to manage our times. And if we use it wrongly,
5	of us instead of knowing nothing at all.Even a piece of information can be crucial such as information about kidnapping,abusing and
6	advantages because the users can find and share information and can be tool for business promotion. Firstly, facebook can be a place
7	role as an announcer. Every updated news and information can be seen through the created group. It's too crucial to the extent
8	of getting everyone under one roof in order to do some discussion can be obviously prevented. Thirdly, Facebook also plays the role of
9	, assignments, lectures, quizzes, and course material etc. Facebook can be used for group study by making a group that is only meant
10	update with anything happen but we must realize that Facebook also can be harmful in our life. Furthermore, all of these things are
11	people. In this case, facebook can help to find a new friend and can be share everything with their friend. So, facebook can be a
12	, business, source of information and news. Firstly, Facebook can be used to connect with family, friends, work colleague and to
13	hours have passed in such a short period of time that you think.You can be infected with Facebook virus. Then almost every minute of
14	if Facebook is used in the right proportions and with proper care, it can be a powerful tool for marketing and networking. Specially, for
15	to avoid this bad things happen. Besides that, Facebook also can be life life threatening sometimes. Many unknown people can
16	read our status and profile, we also can private our profile and this can be save for ourself. Furthermore, students can use facebook as
17	more expensive and student are not afford to pay it. Plus, students can be overthinking about money and they can't perform their best in
18	acid attack. Therefore, facebook should be handled wisely and it can be our best friend as well as our enemy. In fact, we cannot
19	have the advantages. First of all, we can make a private group that can be customize to the certain members. Thus, we can connect
20	do all these thing even we don't who they are. Second,facebook can be used for business.We can start our business by creating one
21	the hard copy. Moreover, cancellation of class and any updates can be posted in facebook. The class representatives do not have to
22	is the web traffic it drives towards your website. Users can be directed towards your product site through links posted on
23	and can be tool for business promotion. Firstly, facebook can be a place for people to find and share information. It is the
24	different countries all over the world. Besides, facebook is also can be use for business. Facebook for business means that we are
25	useful for to get information from the others. Secondly is Facebook can be used to do free marketing. In Facebook, entrepreneurs can

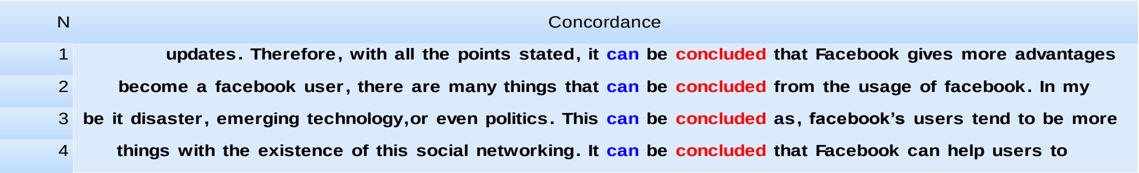
Figure 4.11: Concordance lines for collocate *be*

Table 4.6 also shows shared collocates in MCSAW and LOCNESS measured by the t-score. These demonstrate the certain types of collocates which are found in the two corpora, consisting of pronouns *we*, *you*; preposition/conjunction *from*, *so*; and lexical verbs *make* and *find*. In comparison to the use of these shared collocates in LOCNESS, collocates *we*, *you*, *make*, and *find* were found to be used in roughly the same manner – *we* and *yo*u were basically found to frequently co-occur in the immediate left position of *can*, which suggest the subjective function of the personal pronouns in *we can* and *you can*, whereas *make* and *find* were found to co-occur more frequently to the immediate right of *can*, resulting in the clusters *can make* and *can find* that suggest the expressions of the ability or possibility of making or finding something. This shows similar tendencies found in fictional texts of British English whereby *make* and *find* are among the most recurrent verbs to co-occur with *can* (Mindt, 1995), and in turn, highlights plausible features of conversational speech in novice writing.

Interestingly, collocates *from* and *so* were found to be used differently. In LOCNESS, *from* is seen to appear more frequently in the R2 position of *can* such as in phrases *can conclude/draw/benefit/differ/travel/learn from*, which signals the use of preposition *from* following the immediate *can* + verb phrase. However, it is found that the collocate mostly co-occurs in the R5 position in MCSAW, resulting in long phrases that consist of more than one preposition like *can connect* ***to*** *different people* ***from***, *can learn* ***about*** *new culture* ***from***, *can communicate* ***with*** *different people* ***from***. Collocate *so* on the other, is found to be more frequent in the L3 position in LOCNESS, and L2 position in MCSAW. Most instances in LOCNESS indicate the pattern of ‘so + that + N + *can*’ (e.g. *so that they/he can*). Instances of the pattern in MCSAW however, show more use of the *so* at the beginning of a sentence like ‘***So***, we can say that Facebook is the easier ways for business and entertainment’ (MCSAW_217.txt). This in turn, adds to the more spoken-like feature of Malaysian learner English texts.

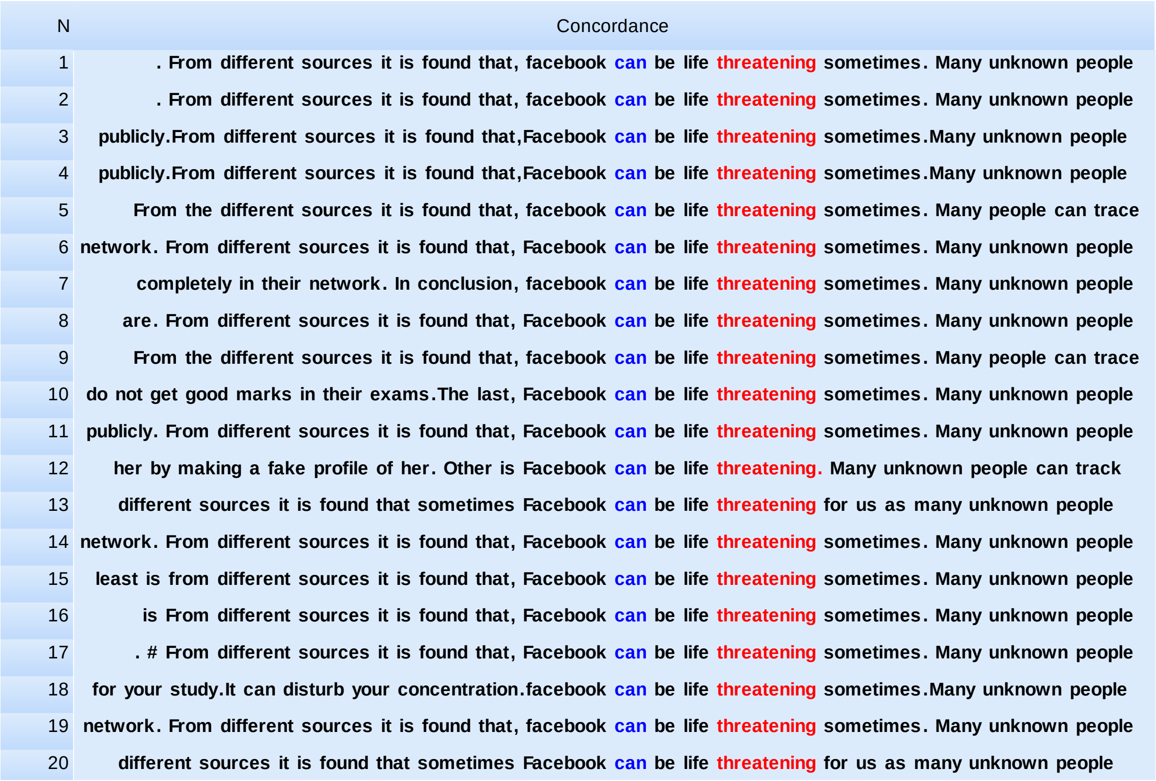
4.2.3.2 Collocates identified using MI-score

The strongest collocates of *can* in MCSAW measured using the MI-score is presented in Table 4.7. They include *concluded* (5.56), *threatening* (5.56), *what's* (5.56), *leverage* (5.56), and *duty* (5.3). It is found that *concluded* co-occurs mostly (4 times) to the R2 position of *can*, resulting in the ‘*can* + be + past participle’ structure, i.e. *can be concluded* as shown in Figure 4.12. Despite the strong association of the collocate with the modal verb *can*, the passive cluster *can* *be concluded* is rarely found in both corpora (occurs only once in LOCNESS). This in turn, implies that the cluster is not found to be a feature in novice writers' argumentative writing. The collocate *threatening*, which mostly co-occurs in the R3 position of *can*, seems to show similar instances across the 20 times it occurs in MCSAW. Similarly, the same is found to happen with each of the remaining collocates – *what's*, *leverage* and *duty*. While each line was checked to ensure they are from separate text files, evidence suggests that learners may over-use these sentences as common examples learnt in the classroom. Otherwise, these similar lines may indicate copying on the students' part.



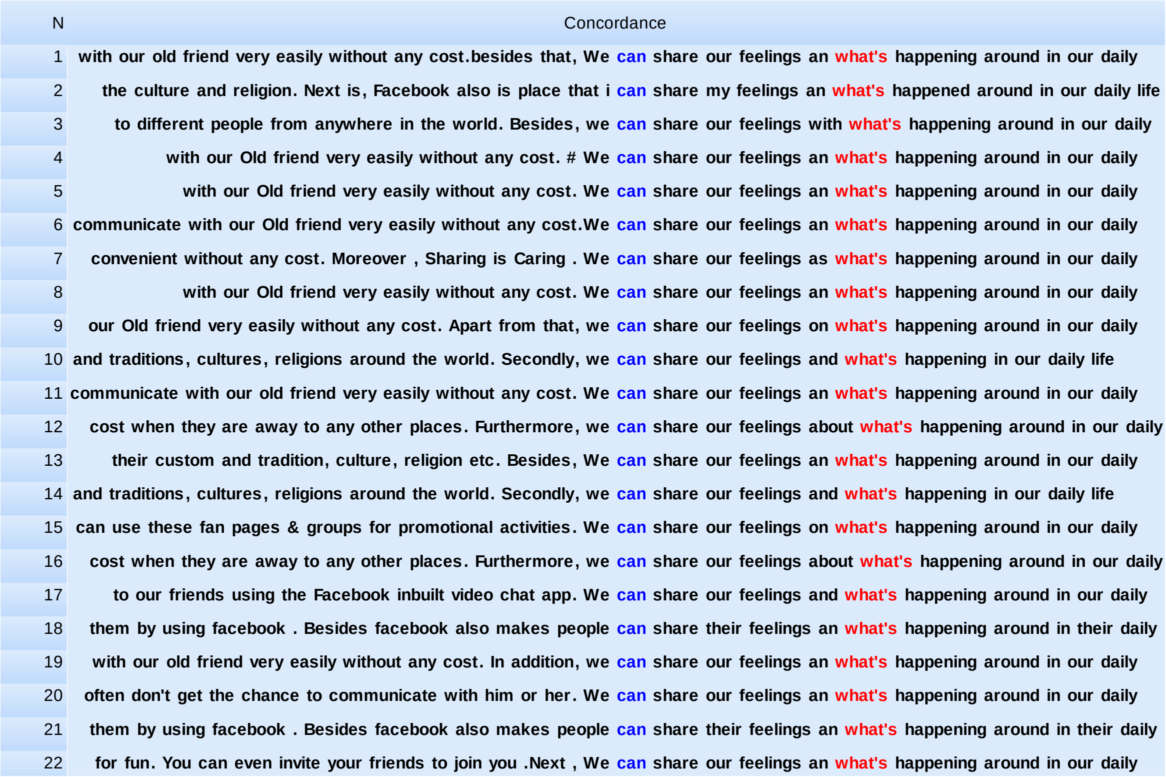
N	Concordance
1	updates. Therefore, with all the points stated, it can be concluded that Facebook gives more advantages
2	become a facebook user, there are many things that can be concluded from the usage of facebook. In my
3	be it disaster, emerging technology,or even politics. This can be concluded as, facebook's users tend to be more
4	things with the existence of this social networking. It can be concluded that Facebook can help users to

Figure 4.12: Concordance lines for collocate *concluded*



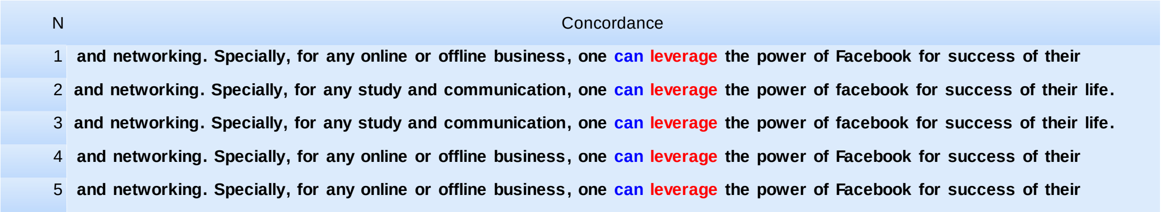
N	Concordance
1	. From different sources it is found that, facebook can be life threatening sometimes. Many unknown people
2	. From different sources it is found that, facebook can be life threatening sometimes. Many unknown people
3	publicly.From different sources it is found that,Facebook can be life threatening sometimes.Many unknown people
4	publicly.From different sources it is found that,Facebook can be life threatening sometimes.Many unknown people
5	From the different sources it is found that, facebook can be life threatening sometimes. Many people can trace
6	network. From different sources it is found that, Facebook can be life threatening sometimes. Many unknown people
7	completely in their network. In conclusion, facebook can be life threatening sometimes. Many unknown people
8	are. From different sources it is found that, Facebook can be life threatening sometimes. Many unknown people
9	From the different sources it is found that, facebook can be life threatening sometimes. Many people can trace
10	do not get good marks in their exams.The last, Facebook can be life threatening sometimes. Many unknown people
11	publicly. From different sources it is found that, Facebook can be life threatening sometimes. Many unknown people
12	her by making a fake profile of her. Other is Facebook can be life threatening. Many unknown people can track
13	different sources it is found that sometimes Facebook can be life threatening for us as many unknown people
14	network. From different sources it is found that, Facebook can be life threatening sometimes. Many unknown people
15	least is from different sources it is found that, Facebook can be life threatening sometimes. Many unknown people
16	is From different sources it is found that, Facebook can be life threatening sometimes. Many unknown people
17	. # From different sources it is found that, Facebook can be life threatening sometimes. Many unknown people
18	for your study.It can disturb your concentration.facebook can be life threatening sometimes.Many unknown people
19	network. From different sources it is found that, facebook can be life threatening sometimes. Many unknown people
20	different sources it is found that sometimes Facebook can be life threatening for us as many unknown people

Figure 4.13: Concordance lines for collocate *threatening*



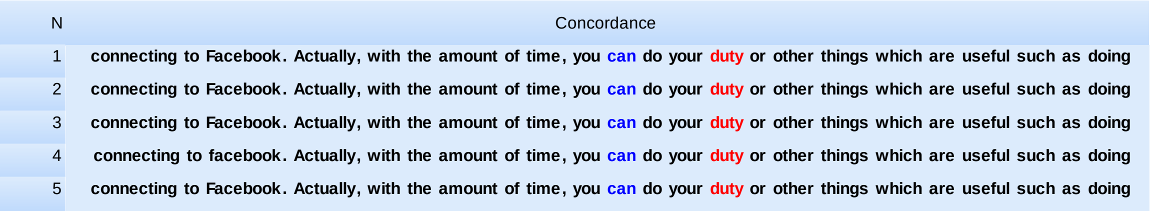
N	Concordance
1	with our old friend very easily without any cost.besides that, We can share our feelings an what's happening around in our daily
2	the culture and religion. Next is, Facebook also is place that i can share my feelings an what's happened around in our daily life
3	to different people from anywhere in the world. Besides, we can share our feelings with what's happening around in our daily
4	with our Old friend very easily without any cost. # We can share our feelings an what's happening around in our daily
5	with our Old friend very easily without any cost. We can share our feelings an what's happening around in our daily
6	communicate with our Old friend very easily without any cost.We can share our feelings an what's happening around in our daily
7	convenient without any cost. Moreover , Sharing is Caring . We can share our feelings as what's happening around in our daily
8	with our Old friend very easily without any cost. We can share our feelings an what's happening around in our daily
9	our Old friend very easily without any cost. Apart from that, we can share our feelings on what's happening around in our daily
10	and traditions, cultures, religions around the world. Secondly, we can share our feelings and what's happening in our daily life
11	communicate with our old friend very easily without any cost. We can share our feelings an what's happening around in our daily
12	cost when they are away to any other places. Furthermore, we can share our feelings about what's happening around in our daily
13	their custom and tradition, culture, religion etc. Besides, We can share our feelings an what's happening around in our daily
14	and traditions, cultures, religions around the world. Secondly, we can share our feelings and what's happening in our daily life
15	can use these fan pages & groups for promotional activities. We can share our feelings on what's happening around in our daily
16	cost when they are away to any other places. Furthermore, we can share our feelings about what's happening around in our daily
17	to our friends using the Facebook inbuilt video chat app. We can share our feelings and what's happening around in our daily
18	them by using facebook . Besides facebook also makes people can share their feelings an what's happening around in their daily
19	with our old friend very easily without any cost. In addition, we can share our feelings an what's happening around in our daily
20	often don't get the chance to communicate with him or her. We can share our feelings an what's happening around in our daily
21	them by using facebook . Besides facebook also makes people can share their feelings an what's happening around in their daily
22	for fun. You can even invite your friends to join you .Next , We can share our feelings an what's happening around in our daily

Figure 4.14: Concordance lines for collocate *what's*



N	Concordance
1	and networking. Specially, for any online or offline business, one can leverage the power of Facebook for success of their
2	and networking. Specially, for any study and communication, one can leverage the power of facebook for success of their life.
3	and networking. Specially, for any study and communication, one can leverage the power of facebook for success of their life.
4	and networking. Specially, for any online or offline business, one can leverage the power of Facebook for success of their
5	and networking. Specially, for any online or offline business, one can leverage the power of Facebook for success of their

Figure 4.15: Concordance lines for collocate *leverage*



N	Concordance
1	connecting to Facebook. Actually, with the amount of time, you can do your duty or other things which are useful such as doing
2	connecting to Facebook. Actually, with the amount of time, you can do your duty or other things which are useful such as doing
3	connecting to Facebook. Actually, with the amount of time, you can do your duty or other things which are useful such as doing
4	connecting to facebook. Actually, with the amount of time, you can do your duty or other things which are useful such as doing
5	connecting to Facebook. Actually, with the amount of time, you can do your duty or other things which are useful such as doing

Figure 4.16: Concordance lines for collocate *duty*

Table 4.7 also shows three shared collocates (*later*, *enjoy*, *found*) measured to be the strongest collocates in both MCSAW and LOCNESS. The collocate *enjoy* was found to be used most frequently in the immediate right of *can* (i.e. *can enjoy*), which is used similarly in both corpora. The instances mainly point towards the subject's ability/possibility of enjoying something. On the other hand, *later* and *found* were used differently in the two corpora. In MCSAW, *later* is mostly used in the L2 position (11 times) such as in *It is very important because* ***later*** *people* ***can*** *sell their products or promote products or services vastly* (MCSAW_437.txt), whereas *later* is found to most frequently co-occur in the R5 position (3 times) in LOCNESS as in *It helps them understand the concept of disciplines, which* ***can*** *be very useful in* ***later*** *life*. Differences between these two types of usage indicate learners' preference to use the adverb *later* immediately following the conjunction *because*. One possible reason is when translated into the Malay language, it is found to signal the use of *later* as a transition marker (*lepas itu*), which generally means *after that*. More specifically, use of *later* in this sense highlights the tendency for Malaysian learners' writing to sound even more spoken-like. Native speakers of LOCNESS on the other, tend to use *later* as an adjective as in the example *later life*. Interestingly, *found* in MCSAW only appears most frequent in a complement clause preceding the main sentence that consists of the modal verb *can* (*it is* ***found*** *that, Facebook* ***can*** *be…*). Furthermore, closer examination of the concordance lines reveals that majority of the 15 most frequent instances of this occurrence are part of the longer repeated sentence that include the collocate *threatening* – *From different sources <u>it is found that</u>, <u>Facebook can be life threatening sometimes</u>* (as shown in Figure 4.13). This could also explain for how they were identified as strong collocates by the MI-score. However, in LOCNESS, *found* is mostly seen to be a part of the cluster *can be found* (8 times) such as in *Computers* ***can be found*** *everywhere from schools to huge businesses* (LOCNESS_USARG.txt), which also occurs in MCSAW (12 times) after the cluster *it is found that*.

To summarise, it can be argued that the modal verb *can* is a preferred marker in MCSAW texts, evenly distributed throughout all parts of the texts, and most frequently co-occurs with pronouns (*we, you*), preposition/conjunction (*from, so*), and high-frequency lexical verbs (*make* and *find*). Even though it was found that both groups of novice writers in MCSAW and LOCNESS produce similar uses of the modal verb *can* (e.g. *we can*), Malaysian writers seem to use longer, more complex prepositional phrases (e.g. *can connect to different people from…*). The use of *so* at the beginning of most sentences in MCSAW further projects a higher tendency for speech written down. In addition, it has been found that *can* strongly associates with several words (e.g. *concluded, threatening, what's, leverage, duty*) and examination of the concordance lines reveal that they are mostly duplicated in more than one text in the learner corpus. This could be evidence for plagiarism within texts, or the overuse of certain prompt sentences taught in the classroom or provided as template prior to the essay production. In the next section, modality is further investigated qualitatively in relation to the modal verb *can*, and how it is used differently by Malaysian learners as compared to the reference language variety in terms of its meanings. In so doing, we are able to understand whether the overuse of modal verb *can* in MCSAW reflects Malaysian learners' style of writing in English or whether it is influenced by other possible factors such as pedagogical implications or influence of L1 transfer.

4.3 Modal meanings for *can*

In this section, qualitative analyses of modal meanings for *can* are investigated via use of concordancing. The English modal system has been studied from various perspectives including Coates (1983) and Palmer (2001, 1990). The terminological, taxonomical and analysing details vary among these works; however, as already mentioned above, "[t]raditionally, the major distinction is between *deontic* and *epistemic modality*" (Krug, 2000: p. 41). Deontic meaning is expressed by linguistic forms that usually indicate obligation and permission. In English, forms like *must*, *should*, *may*, *can*, *permission*, *obliged*, convey deontic modality (Coates, 1990: p. 54). On the other hand, epistemic meaning is expressed by linguistic forms which indicate the speaker's confidence or lack of confidence in the truth of the proposition expressed in the utterance. Lexical items such as *perhaps*, *may*, *must*, *possible*, *I think*, as well as certain prosodic and paralinguistic features, are used in English to express epistemic modality (Coates, 1990: p. 54).

Palmer offers a more detailed model, as shown in Figure 4.17. According to Palmer (2001), modality can be categorised into two major types: Propositional modality and Event modality. Propositional modality is further classified into two types, which are epistemic and evidential modality. The two are distinguished in terms of how a certain proposition is expressed, wherein the latter includes evidence for its claim, while the former does not. In contrast, deontic and dynamic modality are classified under Event modality. Palmer (2001) notes that the difference between the two lies in the conditioning factors which are external in the case of deontic modality, and internal in the case of dynamic modality. This means that deontic modality "relates to obligation or permission, emanating from an external source, whereas dynamic modality relates to ability or willingness, which comes from the individual concerned" (ibid: pp. 9-10).

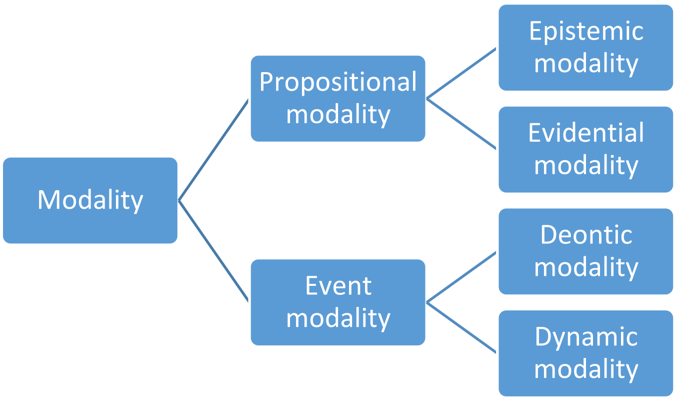


Figure 4.17: Palmer's (2001) classification of modality in modal systems

One way to distinguish between types of modality is through use of paraphrasing. Palmer (2001) explains this in the following examples, in which different categories of modality can be differentiated by the use of 'possible', 'necessary', 'that' and 'for':

(1) Kate *may* be at home now [It is possible (possibly the case) **that** Kate is at home now]
(2) Kate *must* be at home now [It is necessarily the case **that** Kate is at home now]

(3) Kate *may* come in now [It is possible **for** Kate to come in now]
(4) Kate *must* come in now [It is necessary **for** Kate to come in now]

Sentences (1) and (2) indicate propositional modality, in which "the speaker's judgment of the proposition that Kate is at home" (Palmer, 2001: pp. 7-8) is understood with the use of 'that'. Sentences (3) and (4) imply "the speaker's attitude towards a potential future event, […] of Kate coming in" (ibid.) given the use of 'for' and is referred to as event modality. With respect to the modal verb *can*, Palmer (2001) suggests that the modal verb *can* may be equally deontic or dynamic depending on the situation. Table 4.8 below shows the different types of deontic and dynamic *can* according to Palmer (2001) and their examples.

Table 4.8: Types of deontic and dynamic *can* adapted from Palmer (2001)

Deontic	Directives (1) You *can* go now (Paraphrase: You are permitted/allowed to go now) Subjectivity (2) You *can* smoke in here (Paraphrase: It is possible for you to smoke in here)
Dynamic	Ability and willingness (3) My destiny's in my control. I *can* make or break my life myself (Paraphrase: My destiny's in my control. I am able to make or break my life myself) (4) He *can* run a mile in under four minutes (Paraphrase: He is able to run a mile in under four minutes)

In contrast to Palmer (2001), 'Subjectivity' is regarded by Coates (1990, 1983) as a category in itself – termed 'Possibility'. According to Coates (1983: p. 85), the modal auxiliary *can* can be examined in terms of three meanings: 'Permission', 'Possibility', and 'Ability'. She asserts that *can* mainly denotes the two cores – 'Permission' and 'Ability', while 'Possibility' is assigned as an unmarked meaning (Coates, 1990, 1983). This is because there is a continuum of meaning extended from the core meanings of Permission (deontic) and Ability (dynamic) to the periphery of Possibility, which she identifies via "gradients of restriction and inherency" (Coates, 1990: pp. 57-58), respectively. Similarly, this is argued by Imran Ho (1993) as 'dynamic possibility', in which there is an area of overlap between meanings of 'Ability' and 'Permission'. More specifically, Coates (1983) argues that there are indeterminate cases in which it is difficult to decide whether the property in question is determined by external or internal conditions, and thus asserts that "where there is no clear indication either of restriction or of inherent properties of the subject, then 'Possibility' is the meaning which applies" (ibid: p. 93). The distinctions involved are shown below:

> I can do it — Permission = human authority/rules and regulations [i.e. sense
> of restriction] allow me to do it
> I can do it — Possibility = external circumstances allow me to do it
> I can do it — Ability = inherent properties allow me to do it
>
> (Coates, 1990: pp. 57-58)

It is also important to add that where *can* denotes the possibility meaning, it occurs with present/future time orientation (or timelessness) and is followed by the bare/passive infinitive (Mindt, 1995: p. 74). On the other hand, in cases where the modal verb indicates the ability meaning, it occurs with present time orientation or timelessness and is followed by the bare infinitive (ibid).

In short, research has shown that the modal auxiliary *can* signifies two types of core meanings: deontic and dynamic modality. This means that *can* normally conveys 'Permission/Directives' meaning where external conditions (i.e. sense of restriction) are evident, and *can* denotes 'Ability/Willingness' meaning when the possibility of the action is determined by inherent qualities of the subject (i.e. internal conditions). In distinguishing between the two, utterances can be paraphrased with use of *permitted, allowed,* and *able to*. However, where indeterminate cases are found and distinctions are difficult to be made, these can be paraphrased as *it is possible for…* and thus, convey the 'Possibility/Subjectivity' meaning instead. In order to further analyse the frequent modal verb *can* in MCSAW, an adaptation of Palmer's (2001) and Coates' (1983) description and identification of modality is adopted and further elaborated in the following section.

4.3.1 Categorisation of modal verb *can*

Analyses of *can* occurrences in the Malaysian corpus are categorised in terms of the three broad headings, i.e. 'Permission/Directives', 'Ability/Willingness', and 'Possibility/Subjectivity'. Table 4.9 illustrates the criteria and examples following the process of categorising *can* for means of further qualitative analysis.

For *can* functioning as '**Permission/Directives**' (Category 1), instances are understood as acts of seeking/granting permission, which are indicative of core Deontic modality. As previously discussed, use of *can* meaning permission may be identified through internal/external factors that make the particular action possible or impossible, depending on circumstance. In order to determine between the two, use of paraphrases are made with *allow* or *permit* that signals 'Permission'. For example, *You can go now* can be paraphrased as *You are permitted /allowed to go now*. It is noted in Coates (1983: p. 88) that "there is no non-arbitrary way to draw the line" between the internal/external factors thus, for the analysis of *can* in MCSAW, ‘Permission’ meanings that are influenced by subjective factors not found in the context (i.e. subjective deontic modality) are grouped under the ‘Possibility’ category.

In determining the second category, examples of *can* that indicate the ‘**Ability/Willingness**’ meaning or core Dynamic modality are examined in terms of the subject’s capacity or skill to do something. For instance, *He can run a mile in under four minutes* refers to the subject’s physical ability to run within a specific time frame and can be paraphrased as *He is able to run a mile in under four minutes*. Coates (1983) also states that *can* indicating the ‘Ability’ meaning can contain verbs of perception: *see*, *hear, feel*, etc. – often to be found in spoken English. However, Palmer (1990: p. 85) asserts that subject orientation is also “possible with inanimate [subjects], where it indicates that they have the necessary qualities or ‘power’ to cause the events to take place”. This includes instances like *The plane has a built-in stereo tape recorder which can play for the whole four hours*, which suggests that the inanimate subject (the plane’s built-in stereo tape recorder) has the ability to play for the whole four hours. Following this, Palmer’s addition to determining subject orientation is also adapted. In cases where it is difficult to decide whether *can* (‘Ability’) refers to an inherent capability of the particular subject (animate or inanimate) or not, the respective example is categorised under Category (3) ‘Possibility’. This is due to the possibility of the action as determined by “a combination of the inherent properties of the subject and of external factors” Coates (1983: p. 93). As a result, the classification of such instances as meaning ‘Possibility’ is preferred.

Finally, where instances are too ambiguous or do not fit the criteria mentioned in categories (1) and (2) above, they are grouped under Category (3) ‘**Possibility/Subjectivity**’. As described in both Palmer (2001) and Coates (1990, 1983), ‘Possibility/Subjectivity’ meanings can be differentiated by the use of paraphrase ‘it is possible for…’ and ‘it is possible that…’ It is also necessary to point out that use of the paraphrase ‘it is possible for’ tends to refer to Event modality (e.g. deontic/dynamic), while ‘it is possible that’ tends to mean Propositional modality (e.g. epistemic/evidential). In addition, subjectivity is also referred to “words and phrases which are used by speakers of English to qualify their commitment to the truth of the proposition expressed in their utterance”, such as *perhaps*, *I think/believe* (Coates, 1987: p. 112). The next section presents the qualitative analysis for meanings of *can* in MCSAW with respect to the categorisation adopted and adapted from Palmer (2001) and Coates (1983).

Table 4.9: Categorisation of modal verb *can*

Category	Definition	Examples
Permission/Directives	instances that indicate something/someone to be allowed to do something or to have the right or power to do something	*You can go now* (Use of paraphrase with *You are permitted /allowed to*)
Ability/Willingness	instances that have not taken place but are merely potential depending on the subject's intention or desire	Animate subject – e.g. *He can run a mile in under four minutes* (Paraphrase: *He is able to run a mile in under four minutes*) Inanimate Subject Orientation – e.g. *The plane has a built-in stereo tape recorder which can play for the whole four hours.* (Paraphrase: *The tape recorder has the ability to/is able to play for the whole four hours.*)
Possibility/Subjectivity	when instances are too ambiguous or do not fit the other criteria mentioned in 'Permission' or 'Ability'. 'Possibility' meanings can be differentiated by use of paraphrasing – 'possible for' (event modality) and 'possible that' (propositional modality)	*It is possible for* is used to indicate deontic modality – e.g. *You can smoke in here* (Paraphrase: *It is possible for you to smoke in here*) *It is possible that* can be paraphrased with 'perhaps' or 'I think/I believe' to indicate epistemic modality – e.g. *Rain can happen at any minute now.* (Paraphrase: *It is possible (possibly the case) that rain will happen any minute now/ I believe rain will happen any minute*)

4.3.2 Meanings of *can* in MCSAW

To investigate the many types of *can* identified in the Malaysian corpus, 10% of the total *can* occurrences (4,178) were randomly selected using WordSmith's 'random thinning' function. Following this, each concordance line for the 418 *can* instances was examined in terms of the categorisation of modal verb *can* as described in Section 4.3.1. Figure 4.18 presents the types of modality of *can* found in the 418 instances.

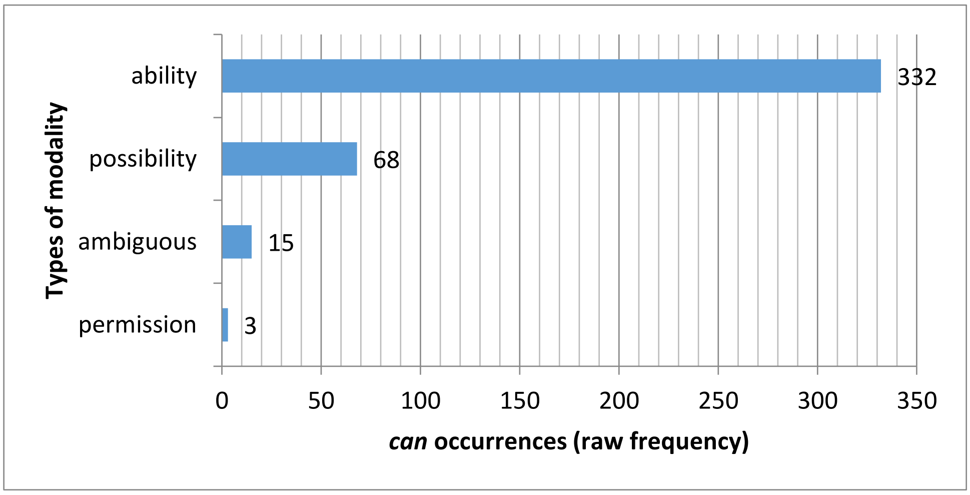


Figure 4.18: Types of modality in the 418 random selection of *can* instances in MCSAW

As can be seen, the most frequent type of *can* used in MCSAW is characterised in the 'Ability' category (79%). The other three categories are found to occur less than 20% each ('Possibility' 16%, and 'Permission' 1%). This initial observation concurs with past findings from Mohamed Ismail et al. (2013) that *can* is mostly used in MCSAW to express a sense of ability than other functions of modality. It is found that few instances of *can* in the learner corpus function as stating permission (3 instances), while there are some examples that are categorised as conveying possibility (68 instances). Also, 4% of *can* examples are listed as ambiguous because of erroneous grammatical sentences that make it hard to determine the meaning of these examples, including:

(5) First of all, the advantages of Facebook to us are **can** find many friends in social networking. (306.txt)
(The subject of sentence is hard to determine: Facebook or we?)

(6) Thus, it is undeniable that Facebook **can** associate with people is one of the advantages. (334.txt)
(This sentence is grammatically incorrect and the subject of *can* is unclear)

Ambiguous examples are therefore discarded and not analysed further. The following parts of this section in turn, are discussion with regard to the three meanings of *can* as discussed earlier in Table 4.9, beginning with *can* referring to 'Permission/Directives', then as 'Ability/Willingness', followed by *can* meaning 'Possibility/Subjectivity'.

4.3.2.1 *Can* meaning 'Permission/Directives' (Deontic modality)

In the learner corpus, *can* meaning 'Permission/Directives' is found to occur the least of the three types of modality (1%), occurring only three times in MCSAW. Examples (7), (8), and (9) demonstrate the 'Permission' meanings that are entailed based on restrictions that render an event to take place. This includes the use of Facebook that allows one to contact friends in line (7), the accessibility to Facebook by having internet connection in line (8), and the opportunity to play online games via Facebook in line (9).

(7) Moreover, it also ***can*** *let us* to contact our friend and know how are they (1040.txt)
(Permission – Moreover, it also allows us to contact our friends and …)

(8) Meanwhile, Facebook ***can*** *access* in many places as long as you have internet connection (96.txt)
(Permission – Many places allow access to Facebook)

(9) For someone who loves to play online games, Facebook is one of the places that they ***can*** *do so* (43.txt)
(Permission – Facebook permits us to play online games)

It should be noted that while there are no examples of *can* denoting the 'Permission' core (i.e. directives); the instances above are characteristic of the deontic meaning for expressing permissibility. More specifically, it signals that, in the writer's judgement, events can only take place through factors that allow the event to be realised (e.g. use of Facebook). According to Imran Ho (1993), the Malay equivalent of *can*, i.e. *boleh* is often used this way to show expressed permission as in examples (7), (8), and (9). He also adds that *boleh* does not necessarily point towards whether the speaker directly allows/permits the subject to do something, but also "where the speaker considers the action/event to be 'the right thing to do'" (Imran Ho, 1993: p. 39). It should also be noted that alternative paraphrases are possible at least for examples (8) and (9) which point to a Possibility/Subjectivity meaning ('it is possible to access Facebook in many places; 'one of the places that it is possible to do so'), but which would not contradict the finding that Permission uses of *can* are rare in MCSAW.

4.3.2.2 *Can* meaning 'Ability/Willingness' (Dynamic modality)

The most frequent meaning of *can* found in MCSAW conveys the dynamic sense (79%). This type of meaning is subject-oriented, in which *can* expresses the subject's ability to perform an action. Furthermore, as mentioned in Mindt (1995), it is usually followed by the bare infinitive. For example, *Students* ***can*** *use Facebook for group study by creating group only for studying* (177.txt). In this sentence, the ability to use Facebook as a group study refers to the subject (Students), in which the modal verb *can* is followed by a bare infinitive (*use*). Unlike *can* meaning 'Permission/Directives' (i.e. Deontic modality), *can* meaning 'Ability' is regarded as a category *internal* to the event taking place (Palmer, 1990). These include the 332 examples that refer to the ability, respectively the willingness of the subject to carry out action denoted by the main verb. Also, the possibility of the action is determined by inherent properties of the subject, as in the innate/intrinsic characteristics of the animate/inanimate subjects of these sentences. Similarly, this includes subject orientation that refer to inanimate subjects having "the necessary qualities or 'power' to cause the events to take place" (Palmer, 1990: p. 85).

In MCSAW, there are 202 examples of *can* instances with animate subjects in MCSAW. These include different types of people (e.g. *friends, hackers, members, students*). As mentioned earlier, pronouns such as *they*, *we* and *you* are seen to be prevalent as reference to this group of people and thus, are significantly found with *can*. Verbs that co-occur with *can* in these examples have been proven to be certain collocates of the modal verb *can* (as discussed in Section 4.2.3) and they mostly refer to actions (e.g. *use, chat, play, create, change*). Furthermore, examples (10) through (14) demonstrate the use of *can* that shows the realisation of each proposition given the internal attributes of the respective subjects:

(10) You ***can*** use these Fan pages and groups for promotional activities. (11.txt)

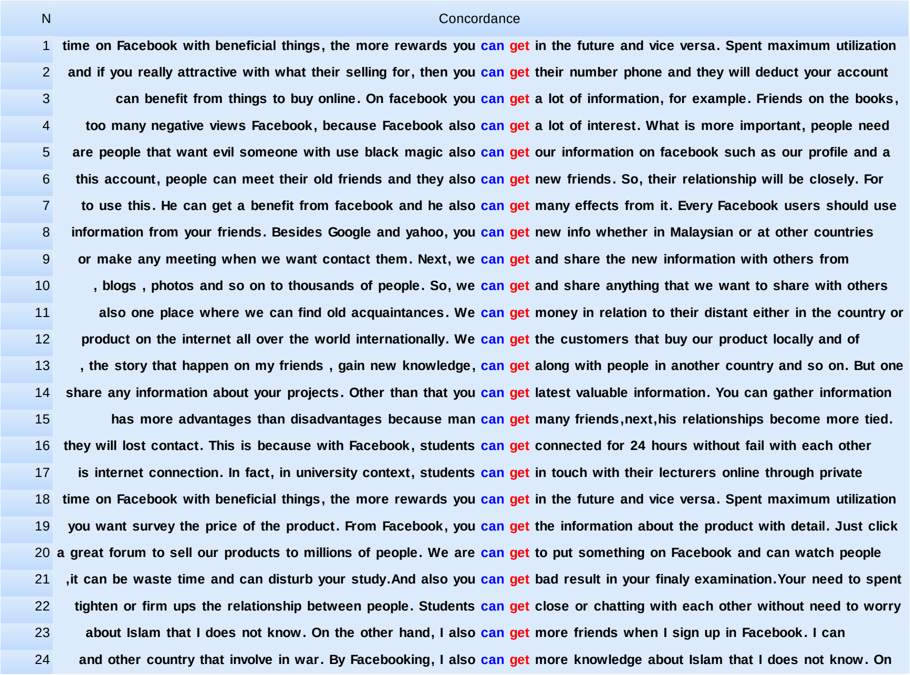
(11) It's a place where we ***can*** chat with others, share our ideas, ask questions, comment on people's status, update our status, make friends, market our business, advertise our products and much more. (305.txt)

(12) We ***can*** play the games when we are free to relax ourself. (1065.txt)

(13) Users ***can*** create profiles with photos, lists of personal interests, contact information, and other personal information (376.txt)

(14) For example, students ***can*** change their opinion and idea about their task,tutorial,paper work and assignment with their friends (315.txt)

It is interesting to note that *can* expresses the 'Ability' meaning in majority of the phrase *can get* (247 times) similar to how it would be translated in the Malay language *boleh dapat*. This phrase in Malay culturally entails that the speaker wishes to convey the ability to attain something, which is usually certain or indicating high probability. Figure 4.19 presents some concordance lines from MCSAW that illustrate the highly frequent use of *can get*. While there is no direct evidence for this inference, it is possible to tentatively hypothesise that the national slogan 'Malaysia Boleh!' might have influenced learners: Over the past two decades, the slogan, which basically means that 'Malaysia can do it'; has been used throughout the country mainly to instil patriotism and inspiration via mainstream media. In turn, learners might be inclined to positively demonstrate their expression for something that is doable or achievable in many parts of their argumentation. Although in some cases the paraphrase *possible for* is also available, these examples occur with animate subjects and therefore are classified as 'Ability' rather than 'Possibility'.



N	Concordance
1	time on Facebook with beneficial things, the more rewards you can get in the future and vice versa. Spent maximum utilization
2	and if you really attractive with what their selling for, then you can get their number phone and they will deduct your account
3	can benefit from things to buy online. On facebook you can get a lot of information, for example. Friends on the books,
4	too many negative views Facebook, because Facebook also can get a lot of interest. What is more important, people need
5	are people that want evil someone with use black magic also can get our information on facebook such as our profile and a
6	this account, people can meet their old friends and they also can get new friends. So, their relationship will be closely. For
7	to use this. He can get a benefit from facebook and he also can get many effects from it. Every Facebook users should use
8	information from your friends. Besides Google and yahoo, you can get new info whether in Malaysian or at other countries
9	or make any meeting when we want contact them. Next, we can get and share the new information with others from
10	, blogs , photos and so on to thousands of people. So, we can get and share anything that we want to share with others
11	also one place where we can find old acquaintances. We can get money in relation to their distant either in the country or
12	product on the internet all over the world internationally. We can get the customers that buy our product locally and of
13	, the story that happen on my friends , gain new knowledge, can get along with people in another country and so on. But one
14	share any information about your projects. Other than that you can get latest valuable information. You can gather information
15	has more advantages than disadvantages because man can get many friends,next,his relationships become more tied.
16	they will lost contact. This is because with Facebook, students can get connected for 24 hours without fail with each other
17	is internet connection. In fact, in university context, students can get in touch with their lecturers online through private
18	time on Facebook with beneficial things, the more rewards you can get in the future and vice versa. Spent maximum utilization
19	you want survey the price of the product. From Facebook, you can get the information about the product with detail. Just click
20	a great forum to sell our products to millions of people. We are can get to put something on Facebook and can watch people
21	,it can be waste time and can disturb your study.And also you can get bad result in your finaly examination.Your need to spent
22	tighten or firm ups the relationship between people. Students can get close or chatting with each other without need to worry
23	about Islam that I does not know. On the other hand, I also can get more friends when I sign up in Facebook. I can
24	and other country that involve in war. By Facebooking, I also can get more knowledge about Islam that I does not know. On

Figure 4.19: Concordance lines for *can get*

The remaining examples in MCSAW show the use of *can* meaning 'Ability' in reference to inanimate subjects, particularly those that refer to Facebook (130 times). These include sentences (15) and (16) where the subjects are inanimate,

(15) In conclusion, Facebook ***can*** do as much harm as good to your social life. (27.txt)

(16) This facility ***can*** help man to add his friend as much as he can. (260.txt)

The intrinsic properties and qualities of 'Facebook' in (15) are that which accord it the ability to be harmful. In (16), the ability of the inanimate subject 'this facility' (i.e. Facebook) to enable users to meet friends online is due to the central properties of the subject, which points to the features of the social networking site. "These are 'subject-oriented' in that they involve some property, disposition on the part of whoever or whatever is referred to by the subject" (Imran Ho, 1993: p. 39). In other instances, Coates (1983) argues that there is the "possibility of the action [as] determined by a combination of the inherent properties of the subject and of external factors" (p. 93), and therefore not always possible to tell whether learners intended to use *can* as to show 'Ability' (i.e. dynamic modality) or subjective deontic modality ('Possibility'). In turn, instances where the inherent or central properties of the subject is not clearly evident, they are grouped under the 'Possibility/Subjectivity' meaning, "since it is the inherent properties of the subject […] which most clearly distinguishes them as belonging to the 'Ability' core" (Coates, ibid).

4.3.2.3 *Can* meaning 'Possibility/Subjectivity'

The identification of modal verb *can* meaning 'Possibility/Subjectivity' has the second highest occurrence (16%), with 68 instances in MCSAW. As noted by Coates (1983), in cases where it is difficult to determine the conditioning factors whether they are internal or external to the subject, 'Possibility' is suggested to be more applicable for the meaning of *can*. The following examples do not imply an indication of restriction or inherent properties in which 'Permission' and 'Ability' meanings are mainly characterised, rather they may describe a sense of external circumstances that permits the use of *can* as seen in the examples below.

(17) Hence, it is disheartening to see that Facebook which initially held such promise, turn into something that ***can*** actually impact society in such a negative way. (38.txt)

(18) Although Facebook have many pros that ***can*** benefit us but it also have its cons. (245.txt)

(19) So, student don't need to pay rental every month that ***can*** burden their parents to pay and family expenses can be reduce. (268.txt)

In (17) it is not certain whether it is the inherent qualities, i.e. abilities of 'Facebook' which create the possibilities for the main predication (giving negative impact to society) or whether it is permissible to use 'Facebook' as a means to achieve the main predication (*It is possible for Facebook to impact society in such a negative way*). Similarly, this possibility of 'Ability' and 'Permission' meanings can be found in (18) and (19). Example (18) can be paraphrased as *it is possible for Facebook's many pros to benefit us…* whereas (19) can be paraphrased as *It is possible for students' rental payments to burden their parents…* In both instances, it is not certain whether the events are realised due to the inherent qualities of the subjects (Facebook/students) or whether it is permissible that they are viewed as a means to achieve the events stated, and therefore, use of *can* in the above examples are classified as meaning 'Possibility/Subjectivity'.

It is interesting to note that while the three sentences above all suggest the meaning of 'dynamic possibility', another similar observation can be seen in the syntactic form of *can*, which is used in the sentences. *That* is often found used as a relative pronoun, preceding *can* in the phrase *that can*, which probably conveys a relative or subordinate clause, usually expressing additional information following the phrase. However, Mindt (1995) states that *can* appears more often in main clauses rather than in subordinate clauses. One possible explanation for this would be another Malay equivalent that signals 'Ability' and 'Possibility': *dapat*, which usually precedes a verb. Although *dapat* can also indicate ability in a non-epistemic sense (dynamic modality), it is different to *boleh* since "[*dapat*] often combines with other auxiliaries…, whereas this is not true for [*boleh*]" (Imran Ho, 1993: p. 42).

Some examples that include the use of *that can* are shown in the examples below. Interestingly, the meaning of possibility can be derived from the translation in Malay. If translated, each instance of *that can* has the same meaning of *yang dapat* +verb, where the verb that follows is usually affixed with the active voice affix *meN* or the passive voice affix *di-* (Imran Ho, 1993: pp. 20-21). In such instances, Imran Ho argues that the affixed items function as main verbs. The examples below in turn, demonstrate how *can* is used in their equivalent Malay translation among writers in MCSAW.

(20) Although Facebook have many pros ***that can benefit*** us but it also have its cons (195.txt)

(Translation: Walaupun Facebook mempunyai (ada) banyak kebaikan ***yang dapat memanfaatkan*** kita, tetapi ia juga mempunyai (ada) keburukannya)
(Although it is possible for Facebook to benefit us, is also has its cons.)

(21) Nowaday, all people around the world like to use technology ***that can make*** people work fast and easy (171.txt)

(Translation: Kini, semua orang di dunia suka menggunakan teknologi ***yang dapat membuatkan*** orang bekerja dengan pantas dan mudah)
(Nowadays, it is possible for technology to make people work fast and easy)

Given the translations, example (20) shows *can* as expressing the possible advantages of Facebook for users through the structure of the modal verb *dapat* with an active voice affixed verb: *dapat memanfaatkan*. In (21), *can* is constructed with the affixed verb *dapat membuatkan* to indicate the chance for people to work faster by using technology. While the expression of 'Possibility' meaning for *can* in these instances is deciphered through its translation in the Malay language, there is reason to argue that learners' L1 (specifically Malay) might be influencing these occurrences.

Also, Mindt (1995) states that *can* expressing subjective deontic modality usually include the modal verb occurring with the passive infinitive. In examples (22) and (23), the subjectivity meaning can be understood again, through use of translation.

(22) So, the money from the business ***can*** be used to increase their income to support their life… (385.txt)

(Translated: Oleh itu, duit daripada bisnes ***boleh*** digunakan untuk menambahkan hasil pendapatan bagi menyara keluara mereka…)

(23) So I surely hope that people ***can*** use the facebook with the proper way (1073.txt)

(Translated: Oleh itu, saya sangat berharap yang orang ***boleh*** menggunakan Facebook dengan cara yang betul.)

In both examples (22) and (23), the learner (who is being specific about the benefits of Facebook) specifies the possible and permitted notion of the events (i.e. 'money to be used to support life' and 'using Facebook the proper way'). *Can* is seen to be treated as the Malay equivalent *boleh* by learners in the process of structuring their argumentation. The same concept of Coates' (1983) gradient of restriction from permission to possibility can be seen to apply to *boleh*, and therefore suggests that most of the *can* instances denoting deontic meaning in MCSAW are subjective in nature. This, in turn, points to Coates' (1990: p. 55) description that subjective meaning refers to "meaning which is speaker-based rather than reference-based" and is thus, categorised under the 'Possibility/Subjectivity' meaning. She further asserts that

> [s]ubjectivity and modality are closely linked in speech. In relaxed conversation, one of the things speakers are doing is expressing themselves. Self-expression, or subjectivity, is encoded by speakers in many ways—lexically, prosodically and paralinguistically—but modal forms appear to be the chief lexical exponents of subjectivity (ibid: p. 55).

As a result, learners' use of *can* similar to the L1 equivalent *boleh* demonstrates a spoken feature of Malaysian learner English writing. Despite having an influence from the Malay language, the use of *can* in expressing 'Subjective Deontic' or 'Dynamic Possibility' meaning may also be attributed to the genre and essay topics. It can be argued that the possibility meanings of *can* are related to some form of opportunities, benefits, and advantages. As mentioned earlier, all three sentences of the translated *that can* signal to situations/events that contribute to the prompt of essay questions (i.e. advantages and disadvantages of Facebook/ living in a hostel). Therefore, its use in the examples above may show how *can* is used in such a way to denote some sense or relation to the topic. In turn, this suggests that there are possible conclusions that can be drawn on the effect of genre or essay topics towards the modal verb *can* in learners' argumentative writing. As mentioned by Hyland (1990: p. 69), "the argumentative essay is defined by its purpose which is to persuade the reader of the correctness of a central statement". Findings therefore, suggest that the highly prevalent use of *can* in MCSAW indicate learners' argumentative style of writing as partially dependent on the essay topic, which is in reference to the Malay equivalent of *can* that is *boleh*, and in some cases *dapat*.

4.4 Summary

This chapter explored keywords in the Malaysian learner corpus, combining this with in-depth analysis of the use of the modal verb *can*, which was found to be prevalent in MCSAW. Keywords were firstly examined, which identified several interesting findings. Among the highly significant keywords are topic-related lexical words such as *Facebook*, *information*, *social*, *networking*, *connect*, and *users*. These words point to the aboutness of the corpus in which the top keyword for MCSAW is inarguably a result of the limited topics available in the Malaysian corpus relative to the reference language variety. Findings of functional keywords on the other, reveal differences of writing style between the two corpora. It was found that Malaysian learners produce more adverbs, conjunctions, and discourse markers (*properly*, *carefully*, *almost*, *furthermore*, *moreover*, *especially*, *too*, *nowadays*, *sometimes*, *anytime*, *anywhere*, *foremost*, *actually*, *already*, *or*, *and*, *addition*, *also*, *besides*, *like*, *than*, *firstly*, *secondly*, *lastly*, *thirdly*, *so*, *for*, *because*), pronouns (*we*, *us*, *our*, *your*, *their*, *you*, *them*, *it*, *everyone*, *anything*, *someone*), and prepositions (*with*, *around*, *through*, *beside*, *from*, *among*, *via*, *for*, *than*, *without*) than were found in the reference language variety.

Moreover, contractions illustrate the spoken-like nature of learner writing such as *it's*, and *don't* that were found in MCSAW while common articles *a*, *an*, and *the* were not. Although Malaysian learners' first language has been argued to not consist of articles (Mukundan et al., 2012), most of these findings suggest learners' tendency towards speech written down such as adverbs expressing place and time (e.g. *almost*, *nowadays*, *sometimes*, *anytime*, *anywhere*, and *already*) and indefinite, first and second personal pronouns (e.g. *everyone*, *anything*, *we*, and *you*) as described in Granger and Rayson (1998), and Gilquin and Paquot (2008). Essentially, the keywords analysis resulted in the identification of highly frequent words *can* and *we*, which ranked second and third after *Facebook*. In relation to past scholarship that investigates modality in learner writing (e.g. Chinese, Greek and Japanese in Gabrielatos & McEnery, 2005; German in Romer, 2004), modal use presents numerous challenges to learners given the modal's polysemous attributes as well as alternative ways in expressing modality (i.e. lexical verbs). However, it is worth restating that the analysis shows the overuse of the modal verb *can* in general (through the keyness analysis), but that future research needs to determine if particular types of modality are over-used.

For the purpose of this chapter, *can* was firstly investigated in terms of range, dispersion as well as collocation comparison by using WordSmith Tools and the reference corpus, LOCNESS. Further qualitative analysis was carried out to uncover the various dissimilarities by examining concordance lines on use of the modal verb *can*, following meanings suggested in Coates (1983) and Palmer (2001). It was found that to a certain extent, learners overgeneralise *can*. Furthermore, results suggest that Malaysian learners of MCSAW show a tendency to use *can* that is similar to the Malay equivalent *boleh* and *dapat*. This overuse of *can* with regard to its similar fashion in Malay may be argued for the common practices in constructing argumentative writing in the classrooms in which verbs *boleh* and *dapat* are familiar features of persuasive writing. Overall, the present chapter has argued that while some similarities can be found as regards the use of modal verbs in MCSAW and LOCNESS, *can* was found to be strikingly more frequent among Malaysian learners as opposed to their native speaker counterpart. In fact, *can* was found to be highly distributed across all texts in MCSAW, in contrast to the use of other modals. In the following chapter, similar investigation is conducted on the use of salient personal pronoun *we* in both corpora.

Chapter 5: Pronouns and the Keyword *we*

5.1 Introduction

This chapter continues with the analysis of results for keyword analysis of the Malaysian corpus (MCSAW) against its reference language variety (LOCNESS). As mentioned in the previous chapter, keywords analysis indicates two significant findings, modality (focus on *can*, which was explored in Chapter 4) and pronouns (specifically, *we*) used differently by Malaysian learners than by their native speaker counterparts. In the present chapter, the examination of keyness value, range and collocates is continued with pronouns (in Section 5.2), which ultimately reveals significant insight into the keyword *we* to be analysed. Further qualitative analysis of *we* in MCSAW is then carried out, following past research on the discourse functions of the first-person plural pronoun, which is presented in Section 5.3.

5.2 Pronouns

Apart from modality, another feature of learner writing is the tendency to use first-person pronouns, which illustrates learners' writing as being more expressive and less formal (Paquot et al., 2013: p. 385). Studies based on learner corpora have shown that learners find it problematic to use a stylistically appropriate tone in their writing, and that a comparative analysis of learner data through written and spoken corpora reveals "a strong tendency among learners, regardless of mother tongue, to use spoken-like features in their written production" (Gilquin & Paquot, 2008: p. 45). Such features include the use of first-person pronouns (e.g. McCrostie, 2008; Gilquin & Paquot, 2008). However, Luzón (2009: p. 193) states that "first person pronouns are part of many phraseological patterns strategically used by expert writers to perform rhetorical functions in academic and professional genres". More specifically, expert writers use first-person pronouns to construct their authorial identities as competent and knowledgeable members of a community. She suggests that "since undergraduate students are novice members of the community, it is difficult for them to grasp such generic conventions and to use language accordingly" (Luzón, 2009: p. 203).

In the investigation of pronouns in another type of learner writing, McCrostie (2008) examined the degree of writer presence in English argumentative academic essays written by a group of Japanese EFL learners. His study replicated an earlier study by Petch-Tyson (1998), which aimed to re-evaluate her hypothesis that learner writing resembles speech written down. His findings supported the earlier study, claiming that Japanese learners' writing contains more writer/reader visibility features,[49] particularly first- and second-person pronouns, than native English speaker writing. It was found that Japanese learners used personal pronouns often with mental/cognitive verbs *think* or *believe* to state a personal view or opinion (e.g. *I think*/*believe*) (McCrostie, 2008: p. 110), compared to the native-speaking writers, who tended to use personal pronouns in guiding the reader through the essay. In light of this and similar research, as well as keyness findings as discussed in Section 4.1 of the previous chapter, the next section will begin by investigating the use of pronouns in MCSAW, and whether it is also the case that Malaysian learners produce more writer/reader visibility features as described above.

Table 5.1 shows normalised frequencies of first- and second-person pronouns,[50] following McCrostie (2008), to investigate the use of pronouns in both MCSAW and LOCNESS. Findings for Japanese first- and second-year students are also included for purposes of comparison. McCrostie (2008) is chosen here for comparison because his data also consists of argumentative essays, whereas other studies on learner language that investigate pronouns focus on other types of writing (e.g. Hyland, 2001; Kuo, 1999). Note that McCrostie (2008) does not provide analysis for third-person pronouns in his study; hence, Table 5.1 only presents results for first- and second-person pronouns in MCSAW and the reference language varieties. As can be seen in Table 5.1, the total for first- and second-person pronouns per 50,000 words[51] is lowest in LOCNESS (575), which indicates that the native-speaker counterparts use a lower amount of personal pronouns compared to that of Malaysian and Japanese students. In contrast, Malaysian learners produced the highest number of first/second person pronouns (2,326), compared to both Japanese first- and second-year students, respectively (2,045; 1,155). Interestingly, there are fewer occurrences of first-person singular pronouns, i.e. *I* and *me*, in MCSAW compared to LOCNESS and Japanese first-year essays. Instead, first-person plural pronouns (e.g. *we*) are over-used in MCSAW, with a proportion almost four times more than in LOCNESS, and three times more than in first-year Japanese students' essays. This appears

[49] Features of writer/reader (W/R) visibility are used "to express personal feelings and attitudes and to interact with readers" (Petch-Tyson, 1998: p. 108) and include first- and second- person pronouns, mental process verbs, emphatic particles, evaluative modifiers, imperatives and questions.

[50] Not all of these are pronouns, some may function as determiners (e.g. *our*). For ease of reference, I will use the term *pronouns* in this chapter to refer to both personal pronouns (*I*, *they*) and personal determiners/possessives (e.g. *my, their*). *Their,* then, reflects the "possessor of some entity" rather than the "participant in some process" (Halliday & Hasan, 1976: p. 45).

[51] This normalisation is used following McCrostie (2008).

to suggest that Malaysian students prefer to use the first-person plural pronoun *we* more than the first-person singular pronouns (e.g. *I*, *me*, *my*, *mine*). Hence, this adds to the importance of exploring the first-person plural pronouns with respect to Malaysian learner English writing in more detail.

Table 5.1: Analysis of 1st and 2nd person pronouns

Feature **Total word count**	**MCSAW** **197,308**	**LOCNESS** **324,019**	**Japanese 1st year** **112,220**	**Japanese 2nd year** **82,194**
1st person singular pronouns ***I, me, my, mine***	1,035	1,342	1,833	805
1st person plural pronouns ***we, us, our, ours***	6,072	1,714	2,080	782
2nd person pronouns ***you, your, yours***	2,072	668	681	310
Total first/second person pronouns	9,179	3,724	4,594	1,897
Total first/second person pronouns per 50,000 words	2,326	575	2,045	1,155

Not all of the pronouns in Table 5.1 are ‘key’ in MSCAW. As shown in the Keywords section in Chapter 4 (Table 4.3 and Table 4.4 and Table A.41 in the Appendix), the top six pronouns that are statistically more significant in MCSAW are *we*, *our*, *us*, *your*, *their*, and *you*; and hence, will be analysed further here. Table 5.2 presents an analysis of the keyness values and range for these six pronouns in the Malaysian corpus relative to their occurrence in LOCNESS. It is found that *we* has the highest keyness value (2,666), followed by the remaining pronouns, ranging from *our* (1327) to *you* (668). This indicates that *we* is significantly more frequent in MCSAW than in the reference language variety. Moreover, it is also key in 84% (429 out of 509) of the total texts in MCSAW. While *your* (1099.69) is also statistically significant in the texts, it is only distributed in less than half (218 out of 509) of the corpus (43%). On the contrary, *their* is mostly widespread, i.e. occurring in 459 texts out of 509 texts (90%) in

MCSAW, but does not seem to have highly significant values compared to the rest of the pronouns (731.96). This is because it reflects the higher use of *their* in LOCNESS, which is less frequent when compared to in MCSAW.

Table 5.2: Keyness measures and range of the top six pronouns in MCSAW

	Keyness	Range
We	2666.04	429
our	1327.45	370
us	1319.98	347
your	1099.69	218
their	731.96	459
you	668.27	230

A closer examination of the distribution of these pronouns in MCSAW and LOCNESS is presented in Table 5.3, which presents normalised frequencies (per 100,000 words) in MCSAW compared to LOCNESS (numbers in brackets indicate raw frequencies). It can be seen that *we* is ranked highest in MCSAW, occurring 1,596 per 100,000 words, followed by *their* (1143), *our* (881), *us* (600), *you* (598), and *your* (449). In contrast, *we* occurs less frequently (404) in LOCNESS when compared to *their*, the latter occurring significantly more frequently (673). Other pronouns in the reference language variety range from 255 (*our*) to 53 (*your*), which contrasts with higher frequencies in MCSAW. In line with the keyness analysis, there seems to be a higher frequency of first and second-person pronouns in MCSAW (especially *we*, *our* and *us*) which corroborates previous studies regarding the higher use of first-person pronouns in learner writing (Gilquin & Paquot, 2007; Luzón, 2009; McCrostie, 2008).

Table 5.3: Relative frequency of pronouns in MCSAW and LOCNESS

	MCSAW	LOCNESS
we	1595.6 (3148)	404.4 (925)
our	881.4 (1739)	255.3 (584)
us	599.6 (1183)	92.7 (212)
your	449.1 (886)	53.3 (122)
their	1142.5 (2254)	673.2 (1540)
you	598.1 (1180)	237.4 (543)

As mentioned earlier, *we* is found to be highly significant in learner writing relative to its occurrence in LOCNESS. *We* is also ranked third in the keywords list (after *Facebook*, and *can*), as discussed in the previous chapter. Given the highly frequent use of *we* compared to other pronouns in MCSAW, the remaining sections of this chapter will focus on in-depth analyses of this first-person plural pronoun.

5.2.1 Range of *we* across all L1 backgrounds

A first examination of *we* occurrences from all essays in MCSAW shows that it is mostly used in texts written by students from the Malay group (86%), followed by Chinese students (13%), and Indian students (1%). Figure 5.1 and Figure 5.2 compare *we* occurrences in the three L1 groups with their distribution across texts in the total corpus. Based on these figures, *we* is considered to be fairly proportionate across all learner groups, similar to results obtained for the modal *can* in Section 4.2.1 of the previous chapter. This again indicates that the frequent occurrences of the pronoun may not be entirely indicative of L1 influence, or alternatively, that they are influenced by all L1s. There is also the possibility that the learners' mother tongue may not have much influence on the overuse of *we* as compared to the role of prompts (or essay topics), as mentioned in the analysis of *can* earlier. It is, however, noteworthy that there is a 4% difference in the use of *we* that are found in Malay texts, and thus, (although a small percentage) it can be said that the occurrence is slightly more associated with Malay users. However, competency rather than L1 background may also play a role, and the influence of these L1s on the overuse of *we* is thus a question for future research.

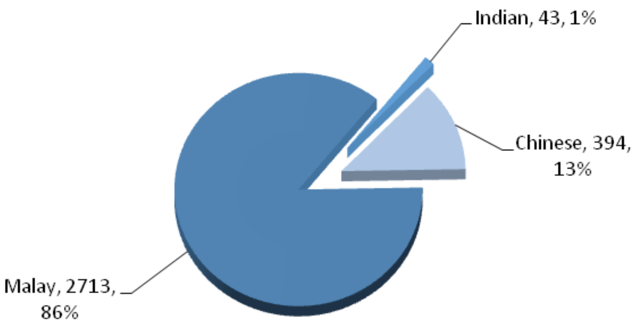


Figure 5.1: *We* occurrences according to L1 groups in MCSAW

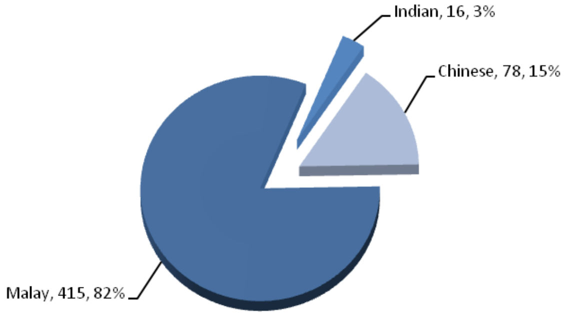
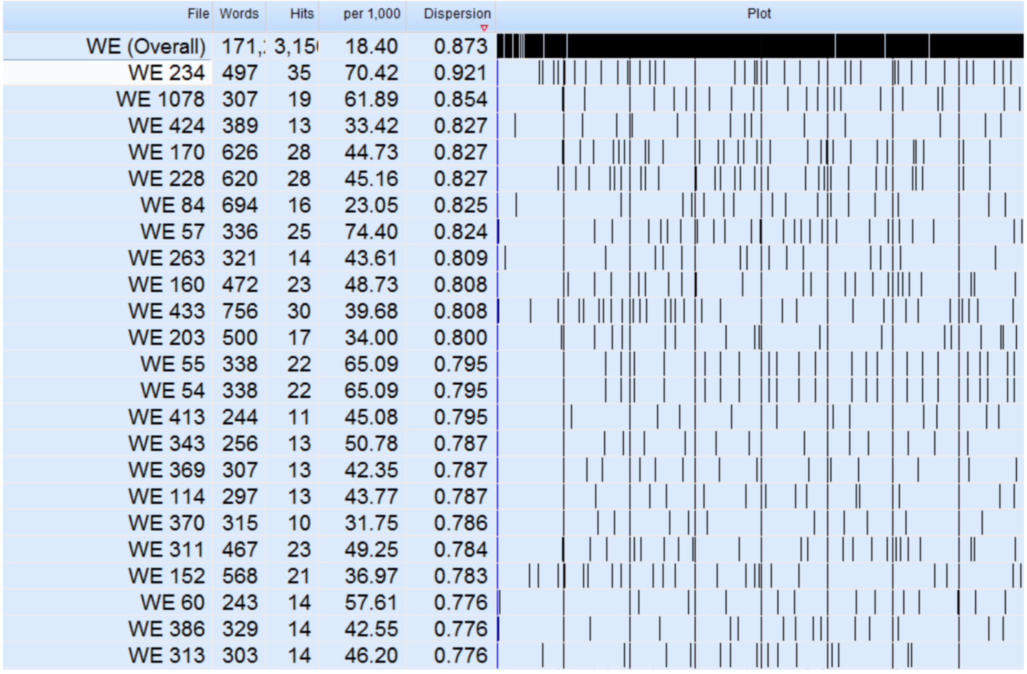


Figure 5.2: Distribution of texts according to L1 groups in MCSAW

5.2.2 Plot and dispersion

The first person plural pronoun *we* is also examined in terms of its plot to see where mention is made most within each text. As mentioned in the previous chapter (Section 4.2.2), investigating the dispersion of selected items promotes the recognition of linguistic patterning that could be representative of a particular genre structure. Similar to *can*, a plot analysis is also conducted for *we* in order to investigate whether *we* occurs in all parts of the texts in MCSAW. Figure 5.3 below presents a sample of the plot diagram illustrating the scattering of *we* across a number of MCSAW texts. Out of the 429 texts in which *we* occurs, 85% had a dispersion value above 0.1. In fact, the first 25 texts, as shown in Figure 5.3, indicate dispersion values close to 0.9, suggesting a very uniform dispersion of *we* in the texts. However, scattering of the pronoun *we* is seen to be less in the beginning part of most of these texts. This means that Malaysian learners do not over-use *we* in the introduction section of their essays, but quite heavily towards the middle and end of their writing.

As mentioned previously (in Chapter 3), Hyland (1990: p. 68) states that the argumentative essay is typically manifested through three stages: Thesis, Argument and Conclusion. Hyland (1990) argues that it is in the argument stage that claims are made in addition to providing support, while the proposition (claim) is once again reinstated in the conclusion along with presentation of its significance. As a result, it can be said that learners' overuse of *we* in the middle and final parts of the essay signals the usage of *we* in the parts concerned with the argumentation and synthesising of claims. Learners' use of the pronoun *we* is thus, further analysed in terms of how they are used in evaluating claims, developing a personal stance, and whether they are used in a coherent manner. In so doing, the following section continues to discuss *we* occurrences in terms of collocational analysis.



File	Words	Hits	per 1,000	Dispersion	Plot
WE (Overall)	171,:	3,15(	18.40	0.873	
WE 234	497	35	70.42	0.921	
WE 1078	307	19	61.89	0.854	
WE 424	389	13	33.42	0.827	
WE 170	626	28	44.73	0.827	
WE 228	620	28	45.16	0.827	
WE 84	694	16	23.05	0.825	
WE 57	336	25	74.40	0.824	
WE 263	321	14	43.61	0.809	
WE 160	472	23	48.73	0.808	
WE 433	756	30	39.68	0.808	
WE 203	500	17	34.00	0.800	
WE 55	338	22	65.09	0.795	
WE 54	338	22	65.09	0.795	
WE 413	244	11	45.08	0.795	
WE 343	256	13	50.78	0.787	
WE 369	307	13	42.35	0.787	
WE 114	297	13	43.77	0.787	
WE 370	315	10	31.75	0.786	
WE 311	467	23	49.25	0.784	
WE 152	568	21	36.97	0.783	
WE 60	243	14	57.61	0.776	
WE 386	329	14	42.55	0.776	
WE 313	303	14	46.20	0.776	

Figure 5.3: Dispersion plot for *we* occurrences in MCSAW

5.2.3 Collocation comparison

Table 5.4 and Table 5.5 show words that most frequently co-occur with *we* in both the Malaysian learner corpus and reference language variety, using t-score and MI. Similar to the process of identifying collocates for *can* in the previous chapter (Section 4.2.3), analyses for *we* collocates are also conducted using both test scores, with a span of 5 words to the right and left, to compare the different results. While results for the MI score reveal more *we* collocates than for the t-score, there are more shared collocates for the personal plural pronoun *we* in terms of both measures, compared to the shared collocates for the modal verb *can* discussed in Chapter 4. This first observation points to the common use of *we* in both corpora, in contrast to *can*, which use is found to be particularly more salient among Malaysian learners than in the reference language variety.

Table 5.4: Collocates for *we* in both MCSAW and LOCNESS using t-score

Only in MCSAW	**Only in LOCNESS**	**In both**
that (13.30), also (12.51), get (12.28), use (12.18), with (11.71), as (11.50), to (11.16), the (10.76), from (10.72), share (10.55), in (10.53), facebook (10.47), a (10.30), have (10.14), not (9.64), friends (9.57), of (9.57), and (9.48), are (9.34), when (9.30), it (9.07), connect (8.95), information (8.69), this (8.43), advantages (8.28), is (8.18), for (8.10), using (8.08), because (8.07), on (7.98), or (7.903), other (7.90), any (7.80), just (7.63)	why (2.39), begin (2.17), perhaps (2.1), come (2.03)	can, our, **know**, if, so, do, **need**, about, should, **find**, **must**, them, **see**, how, what, us, now, could, **ask**, still, **live**, **today**, **say**, then, **ourselves**, **already**

Table 5.5: Collocates for *we* in both MCSAW and LOCNESS using MI score

Only in MCSAW	**Only in LOCNESS**	**In both**
conventional (5.96), tiring (5.96), greetings (5.96), learnt (5.7), wrongly (5.48), mouse (5.38), feedback (5.38), miles (5.33), conclude (5.27), deny (5.2), separate (5.16), sweet (5.16), note (5.12), anybody (5.09), chance (5.09), appreciate (5.06), stand (4.96), allowed (4.96), carefull (4.96), radio (4.96), article (4.91), views (4.85), blogs (4.77), complete (4.7), called (4.68), away (4.67), bored (4.64), correct (4.64), wise (4.64), trust (4.64), couldn't (4.64), put (4.61), needed (4.59), aware (4.57), gather (4.55), freely (4.52), manage (4.52), bully (4.51), clearly (4.45), single (4.43), misuse (4.43), plan (4.38), paid (4.38), trace (4.38), properly (4.30), exchange (4.28), feelings (4.26)	sympathise (6.13), assume (6.06), hear (5.7), willing (5.13), begin (5.1), admire (4.87), stage (4.87), expect (4.81), move (4.68), consider (4.57), hold (4.39), attempt (4.23), let (4.13), animals (4.11), meaning (4.06), perhaps (4.01), accept (3.94), here (3.91), watch (3.78), understand (3.7), responsibility (3.66), process (3.65), ways (3.55), america (3.55), ban (3.5), cars (3.45), come (3.44), longer (3.42), view (3.40), why (3.36), works (3.34), us (3.31), past (3.29), someone (3.28), die (3.28), going (3.27), living (3.26), question (3.25), really (3.24), ever (3.24), sympathy (3.20), throughout (3.11), seen (3.01)	**see**, **must**, **ourselves**, learn, remember, never, don't, **ask**, cultures, enjoy, **need**, **know**, cannot, always, **say**, want, sure, start, realize, **already**, **find**, maybe, **today**, **live**, read, something, said, look, certain

For the purpose of this section, discussion is focussed on the collocates that are identified as collocates in both corpora using the t-score and MI score. These collocates include 11 words, namely *know*, *need*, *find*, *must*, *see*, *ask*, *live*, *today*, *say*, *ourselves*, and *already*, which are highlighted in bold in the above tables. Categorically, these words can be further identified in terms of modality and verbs of necessity/desire (*need*, *must*), mental/cognitive verbs and verbs of discovery/perception (*know*, *find*, *see*), action and speech verbs (*live*, *ask*, *say*), reflexive pronouns (*ourselves*), and adverbs (*today*, *already*). More specifically, it appears at first glance that both sets of novice writers appear to use *we* with similar co-occurring words in their writing.

Firstly, it is common to find *we* co-occurring with expressions of modality, as is shown previously in the keywords analysis of both corpora. However, *we* is found here to be co-occurring more significantly with the words *need* and *must*. Findings reveal that *we* co-occurring with *must* indicates the deontic (necessity) meaning of the modal verb. In turn, most of the examples of *we* + *must* show the intention of both groups of writers in making propositions that involve the reader necessarily doing or acting in a particular way: for instance, in *All the things that we do* ***we must*** *have limit* (MCSAW.243.txt), and *So although life is futile* ***we must*** *live it to the full* (LOCNESS.BRSUR1.txt). Similarly, *we* co-occurring with *need* also refers to the same expression, i.e. a necessity meaning as regard the statement that is being expressed: examples include ***We need*** *to be careful while using Facebook…* (MCSAW_1044.txt) and *…and* **we need** *to shift our focus to a broader view of life rather than focus on the needs of one sick individual* (LOCNESS_USARG.txt). However, another look at the position of *need* co-occurring two words to the right of *we* shows 17 instances of the phrase –*we just need to…*, found to be prevalent in MCSAW, but not found in LOCNESS. Learners' use of the adverb *just* can be compared to Gilquin and Paquot's (2008: p. 46) results for the overuse of the adverb *maybe* in learner writing, which they argue to be more typical of speech than in writing. Furthermore, the adverb *just* is considered extremely common in conversation with "over 2,500 times per million words – far more than any other adverbials in other registers" (Biber et al., 2002: p. 368). Hence, it could be said that the collocational pattern of *we* + 'just need to' in MCSAW also exhibit spoken-like features in writing.

Luzón (2009: p. 201) also found *need* and *must* as collocates of *we* in her study, in which "[*w*]*e* is used with modals and semi-modals as a solidarity strategy intended to involve the reader and build a working relationship" (e.g. *Fujitsu is a well-known brand in electronics, so if we want to ensure our success with our Mobtronic PDA, we must choose Fujitsu…For movement of the train we need permanent magnetic field*). More importantly, these highly associated collocates can be accounted for by "their high use by students to perform specific rhetorical functions" (ibid: p. 196), often used as "a solidarity strategy" with *we* functioning as an inclusive pronoun.[52] This socially defined rhetorical identity is accomplished most visibly in the use of first-person pronouns and possessive determiners (Hyland, 2002a), ultimately to make the reader feel involved (Harwood, 2005b: p. 346).

[52] *Inclusive we* refers to the writer and reader together, whereas *exclusive we* refers solely to the writer and other persons associated with the writer (Harwood, 2005b; Kuo, 1999).

It can also be argued that the ‘solidarity strategy’ is constructed in both corpora using cognitive and speech verbs such as *know*, *find*, *see*, *ask* and *say*. In general, every instance of *we* co-occurring with the before-mentioned verbs is found to function as an inclusive pronoun. This suggests that both groups of novice writers incorporate similar amounts of writer visibility (McCrostie, 2008). Collocates such as *know* reflect Luzón’s findings, that Spanish learners use this verb with the inclusive *we* to relate to their readers, such as in *It is because* ***we know*** *that the market price become increase like rent* […] (354.txt). The same can be found in MCSAW and LOCNESS: *From it,* ***we know*** *that not all of the student has money and come from rich family* (MCSAW_382.txt) and ***We know*** *how people can contract the disease, and* […] (LOCNESS_USARG.txt). It has also been found in Chinese learner writing that *we* is often used with *know* in the phrase ‘as we know’ and ‘as we all know’ (Fawcett, 2013: pp. 259-260). According to Fawcett, “[b]oth of these popular construals restrict the assertion to the knowledge of the in-group through the deployment of the personal we” (2013: p. 259), and thus “situates the text to be inarguable and excludes any reader who may disagree with the assertions” (ibid: p. 305). Such phrases also occur in MCSAW (57 times), as will be discussed further in Section 5.3.2.

However, it appears that the collocate *find* is used differently in relation to *we* in the two corpora. In LOCNESS, *we* is found to co-occur with the verb *find* in R1 position, whereas the verb is seen to occur more frequently in the position of two words to the right of *we* in MCSAW (R2). Interestingly, Malaysian learners use more *we can find* phrases compared to the novice writers of LOCNESS, as shown in *With facebook* ***we can find*** *our long lost friend* (MCSAW_364.txt). 46 examples of this collocation suggest that learners use *we* + *find* to refer to the physical act of finding, with the use of modal verb *can* here expressing the possibility/ability of the action ‘find’. On the other hand, native-speaking writers use instances such as ***We find*** *new meanings to these emotions because they provide a way for us to escape from social expectations* (LOCNESS_USMIXED.txt) to indicate existential meaning.

This is similar to the case with collocate *see*. Malaysian learners are found to use the verb *see* more frequently in the R2 position of *we* (*we can see*, 50 times) compared to seven times occurring immediately to the right of *we* (*we see*). In contrast, *we see* (27 times) occurs more than *we can see* (15 times) in LOCNESS. One possible explanation might be the highly significant use of *can* in learner writing as opposed to the reference language variety. There are also cases in which examples of ‘inclusive *we* + *see*’ mirror those of findings in Fawcett (2013) and Luzón (2009). Luzón states that these typical examples signal existential meaning, i.e. learners tend to use *we see* instead of *there is* (2009: p. 196). This is found in both MCSAW and LOCNESS, in the following instances: *In 1982,* ***we see*** *a turning-point in French industrial relations with the passing of the lois Auroux* (LOCNESS.BRSUR1.txt; paraphrasable as 'there was'), and *As* ***we see*** *now, many husbands and wives are fighting because of Facebook* (MCSAW.181.txt; paraphrasable as 'there are now many husbands ... fighting' or 'many husbands and wives are now fighting'). These collocational phrases are also found to occur with the 'we can see' construction (Fawcett, 2013), that "invoke greater certainty and clarity through perception similar to a more objective form such as 'it is evident' or 'it is certain'" (p. 266). Such examples include *Nowadays,* ***we can see*** *that many social networks in the internet.* (MCSAW_60.txt) and *On studying France throughout the twentieth century* ***we can see*** *that this is indeed true* (LOCNESS_BRSUR.1txt).

For the remaining verbs (*live, ask,* and *say*), both corpora present similar functions for the *we* + action/speech verb collocates. Although there are more instances of *we* co-occurring with *live* in LOCNESS, both corpora use the words in referring to the place where *we live* in: for instance, *We can know around the world even though* ***we live*** *in Malaysia* (MCSAW_117.txt) and *He says that the society* ***we live*** *in today creates an environment* (LOCNESS_USMIXED.txt). Such uses may indicate that both sets of novice writers draw on 'lived experience' in constructing arguments; an experience that is, furthermore, constructed as common to a nation or society – abstracted from the personal to the communal. In contrast, there are more instances of *we* co-occurring with verbs *ask* and *say* in MCSAW, particularly in the R2 position. Such examples interestingly co-occur with the personal pronoun *we* and the modal verb *can* and thus, are used as a sort of hedge, either to emphasise or to minimise a claim, but would be omitted in expert writing (Luzón, 2009). These include examples such as *when we do not know something about our lesson at school,* ***we can ask*** *or post that question at facebook* (MCSAW_54.txt) and ***We can say*** *that everyone opens their account facebook everyday if have free time* (MCSAW_402.txt). Granger (1998) has shown that phrases such as ***we can say that*** are recurring phrases in learner writing, and that they "fill exactly the same function as *actually* or *as a matter of fact*" (ibid: p. 9); and which have also been found to be specifically over-used in French learner writing. Examples in LOCNESS include *But* ***we ought to ask*** *ourselves "What happens when the computer-orientated world collapses?"* (LOCNESS_alevels9.txt) and *When we start the annoucements before the race, the first thing* ***we say*** *is that the gas is on the right and the brake is on your left* (LOCNESS_USARG.txt).

Contrary to the example in learner writing, *we* + *ask* co-occurs with the modal verb *ought*, which expresses the obligation/necessity meaning and is generally rare in comparison with other modals. *We* + *say* can be seen to refer to the action of speech rather than functioning as a hedge as used in learner writing. Nevertheless, they are both visible in the two corpora.

Relative to being a first-person plural pronoun, *we* is also suggested to co-occur with its other forms (i.e. *ourselves*). Examples from both corpora indicate that these other forms are more frequent in the R2 and R3 position compared to occurring immediately after *we*. Moreover, the collocation *we* + *ourselves* is seen to be used similarly by both groups of writers, which is generally seen as reflexive: for example, *with all of the new technology & discoveries* ***we find ourselves*** *struggling to survive the disease AIDS* (LOCNESS_USARG.txt) and *It actually all depends on how* ***we carry ourselves*** (MCSAW_3txt). On the other hand, *we* co-occurring with adverbs (*today* and *already*) can be said to express typical characteristics of the argumentation genre. In both corpora, *today* is used generally to situate the reader to the present context: examples such as *Everywhere* ***we turn today****, the world is crowded with people busy looking for the jobs* (MCSAW_230.txt) and ***We are experiencing today*** *a definite movement towards an ever closer and more integrated Europe* (LOCNESS_BRSUR3.txt) also demonstrate the expression of a 'lived experience', in which novice writers refer to the world or present-day reality in their writing. This consequently can be described as realising the 'solidarity' approach. Similarly, it is also found to be true of *we* co-occurring with *already*, in instances such as ***we already*** *know that Facebook is one of the social networking site in the world* (MCSAW_254.txt) and ***We already*** *know that it is inevitable that when someone becomes infected with HIV* (LOCNESS_USARG.txt). The use of adverb *already*, however, has been shown to be over-used in learner writing and is not frequently found in academic writing (Granger & Rayson, 1998: p. 124). In fact, the phrase 'we already' co-occurs with the mental process verb *know*, and hence, contributes to the high writer/reader visibility in novice texts. Instead, adverbs that express time such as *now, ago, always, often, sometimes, already, still, everywhere, here* are described as more common in spoken discourse (Granger & Rayson, 1998). Most importantly, it can be seen that the overall use of *we* in both corpora show similarity to each other.

So far, frequency results for use of pronouns have supported past research (Cobb, 2003; McCrostie, 2008), in that learners produce higher occurrences of the first-person pronouns than in the reference language variety. As mentioned by Hinkel (2002), first-person pronouns (*we*) signal interpersonal discourse and direct involvement of the writer, and are usually more common in spoken rather than written registers. This, in turn, points to the high writer/reader visibility in learner writing, as argued by Petch-Tyson (1998). Interestingly, results from collocation comparison of *we* indicate a fair share of the use of the personal plural pronoun *we* in both groups of novice writers. This could be due to similar strategies of using *we* in writing argumentative essays, or the comparable level of novice writing. To investigate this hypothesis further, qualitative analysis of the discourse functions of instances that include *we* will be presented in the next section.

5.3 Discourse functions of *we*

This section will move on to qualitative analysis looking at the use of first-person plural pronoun *we*, by use of concordancing. As mentioned earlier, previous studies by McCrostie (2008) and Petch-Tyson (1998) have concluded that non-native English speaker writing contains far more personal involvement than equivalent native English speaker writing, and as a result tends to resemble spoken language. While these two studies have analysed the degree of writer/reader visibility features in the corpora, others have further investigated the discourse functions of first-person personal pronouns (e.g. Kuo, 1999; Luzón, 2009; Tang & John, 1999). More specifically, first-person pronouns are found to be used by academic writers for a range of purposes (Breeze, 2007; Hyland, 2002a). These include,

> stating a purpose or goal, organising the text and making its structure clearer to the reader, staking knowledge claims, stating hypotheses, showing results or findings, expressing personal opinions, conveying a sense of novelty about the author's research, explaining experimental procedures, creating a positive tenor of solidarity, and constructing the author's identity as a member of the discourse community (Luzón, 2009: p. 193).

Hyland (2002b) also emphasises the use of pronouns to construct identity and voice. According to Hyland (2002b: p. 352), "[t]he author's explicit appearance in a text [i.e. by use of pronouns], or its absence, works to create a plausible academic identity, and a voice with which to present an argument". However, learners tend to find this process difficult, partly because of two reasons: constructing these identities can differ considerably from those they are familiar with; or because students are rarely taught that disciplinary conventions differ (ibid). Hyland (2001: p. 557) states that one of the most common use of personal pronouns in academic writing is in the use of the inclusive *we*, in which "[r]eaders are most explicitly brought into the text as discourse participants". First-person pronouns, thus, can act as a rhetorical strategy that allows writers to perform different discourse functions in the text, through which they construct a convincing argument that persuades readers of the validity and novelty of their claims and of their own competence. Some examples of this 'inclusive' use have already been encountered in Section 5.2.3 above.

Further studies have looked into the use of personal pronouns in expert scientific research writing (Kuo, 1999), as well as in general writing and published academic writing (Tang & John, 1999). Both studies found that first-person plural pronouns have a number of semantic references and perform multiple functions in the journal article. They also report that there is a degree of authorial power to the employment of first-person plural pronouns, which is different depending on types of texts. In exploring the many uses of first-person pronouns, Kuo (1999, p. 130) highlights that this involves investigating "the function that a sentence containing a personal pronoun performs in the immediate discourse context". In other words, personal pronouns are examined in terms of how they are used functionally, by the contribution they make to the discourse. This is also the approach taken in the present thesis: that is, I will examine the way the first-person plural pronoun *we* functions together with its co-text.

In so doing, this thesis will partially draw on a study by Luzón (2009), who explored how Spanish EFL Engineering students used the pronoun *we* in a corpus of 55 reports. She found that students used *we* differently to expert writers who use first-person plural pronouns strategically to perform specific discourse functions. Her results show that the Spanish Engineering students produced more first-person pronouns compared to the reference corpus, which confirmed previous findings that there is higher visibility of the author and higher use of spoken language features in learner writing (Neff et al., 2004; Petch-Tyson, 1998).

Luzón (2009) further argues that the Spanish learners are unaware of how expert writers use first-person pronouns to construct their authorial identities as competent and knowledgeable members of a community. By comparing the different types of discourse functions of the personal plural pronoun *we*, Luzón (2009: p. 197) demonstrated how Spanish learners used the pronoun differently to the more conventionalised use of patterns involving *we* to perform specific functions in academic writing. More specifically, she distinguished nine discourse functions of *we*, namely: 1) Stating goals or purposes; 2) Stating conclusions; 3) Expressing a final recommendation; 4) Guiding the reader through the text; 5) Recounting the research process; 6) Showing results or findings; 7) Assuming shared experiences/knowledge, goals, beliefs; 8) Emphasising or calling the reader's attention; and 9) Expressing opinion or volition.

In contrast to most other studies of the discourse functions of *we*, which examine scientific-based and expert or published academic writing, Luzón's (2009) classification scheme of the discourse functions focuses on first-person pronouns in learner writing. This appears to be particularly relevant to the context of examining discourse functions of *we* in the present chapter due to the common features of novice writing. However, it should be borne in mind that her classification scheme is based on learners' report writing, which is not similar to argumentative essays; and therefore, this scheme is adapted for the present chapter. To reiterate, the argumentative essay is the most common genre that undergraduate students have to write (Wingate, 2012), and its main objective is to persuade the reader of the correctness of a central statement (Hyland, 1990).

5.3.1 Categorisation of plural pronoun *we*

Occurrences of *we* in MCSAW are thus classified into categories adapted from Luzón (2009), excluding those that are not related to the argumentative genre (i.e. expressing final recommendation, recounting the research process, and showing results or findings). This is because the texts in MCSAW are not research papers (nor reports), nor are the writers participating as yet in a professional academic discourse community. It is also important to bear in mind that Luzón's (2009) study looked at learners' report writing in comparison to samples of expert academic writing (Kuo, 1999) and journal research articles (Harwood, 2005a, 2005b), all of which are not similar to the genre of argumentative essays in MCSAW. Nevertheless, it is still possible for some of the criteria to be applied to the functions of *we* in the present corpus, where appropriate, as is presented in Table 5.6. In addition, two other readings are also occasionally referred to, when further elaboration is warranted in terms of analysing the results from an argumentative genre perspective (Hyland, 1990; Wingate, 2012).

Table 5.6: Categorisation of plural pronoun *we*

Categories	Definition	Examples
Stating the thesis	this stage introduces the discourse topic or proposition to be argued and advances the writer's proposition or central statement	*There are two advantages* ***we*** *can get…*
Stating conclusions	functions mainly to conclude the essay	***we*** *propose/conclude*
Guiding the reader through the text	statements that signal the different parts of the text and present the content of the subsequent discourse in order to make the structure of the text clear to the reader. It can also be used anaphorically to refer to preceding fragments of discourse	***we*** *continue to discuss*
Assuming shared experiences/knowledge, goals, beliefs	encompasses assumptions of shared experience and some instances of modality (including *can*) and cases where deontic modals or verbs of volition, etc. are not used to express a statement of opinion. Examples are primarily those where *wish*/*want* occur in hypothetical or non-actual contexts.	*we can find* *if we want to; when we wish to*
Emphasising or calling the reader's attention	verbs used to call the reader's attention in expert discourse collocate with adjectives such as *important*, *interesting*, or with the modal verb *should*	***we*** *note that*

Expressing opinion or volition	includes examples where *we* occurs with verbs of necessity (*need*), volition (*want*, *wish*), and deontic modality (*should*, *have to*) as expressions of opinion or volition	***we** need*, ***we** want*

With respect to Luzón's (2009: p. 198) first function of *we*, 'Stating goals or purposes', in research writing this includes statements such as *in this paper **we** report*, often with verbs such as *determine*, *assess*, *address*, *identify*, or *study*, and usually occurs in the present tense or infinitive. According to Hyland (2002a: p. 1100), these functions are mainly "to signal the writers' intentions and provide an overt structure for their texts […], relating to facets of the text which make the organisation of the discourse explicit". However, given the genre of argumentative writing, the related category '**Stating the thesis**' instead of 'Stating goals or purposes' is more appropriate for classification of occurrences of *we* in argumentative essays. Hyland (1990) remarks that there are five possible ways in which the thesis can be realised, but the most common move found in examination scripts (normally argumentative essays) is through use of markers that structure the discourse by signposting its subsequent direction: for example, *There are a number of reasons for increasing assistance to community education*. I will hence, categorise instances of *we* that co-occur with such signposts as 'Stating the thesis'.

On the other hand, *we* used for concluding the essay (**Stating conclusions**) is normally found co-occurring with the most frequent verbs *show*, *conclude*, *demonstrate*, *suggest* and *propose* (e.g. *We* propose/ conclude), and this remains similar for argumentative essays. The difference is that the conclusion relates to the student's argument rather than to the research process or results. Use of *we* for the first two functions (Stating thesis and Stating conclusions) is usually expected in the introduction and conclusion phase of an essay.

The third function (**Guiding the reader through the text**) involves the pronoun *we* used "in statements that signal the different parts/sections of the text and present the content of the subsequent discourse in order to make the structure of the text clear to the reader" (Luzón, 2009: p. 199). For this function, statements with *we* can be found at the end of the introduction as well as in any other part of the text to announce what comes next (either the contents of a section or the contents of the subsequent paragraphs). This particular role is often signalled explicitly by the use of verbs like *see*, *note*, and *observe*, i.e. mental processes of perception, specifically visual perception (Tang & John, 1999: p. 27); and is usually realised in the inclusive form of *we* or *us*. This means that *we* in this sense can be used anaphorically to refer to preceding fragments of discourse.

We also functions in **assuming shared experiences/knowledge, goals, or beliefs** (4th function). Although this is usually constructed as representing a group of people, in many cases "passive voice or sentences with inanimate subjects would be preferred in expert writing" (Luzón, 2009: p. 201). For instance, the sentence *This page is orderly and we can understand it quite well* (Luzón's example) is better rephrased as *This page is orderly and easy to understand*. In cases where the inclusive *we* is used to express shared or common knowledge, expert writers tend to use other devices that make the author less visible. For example, *we can find* and *we have* could be replaced by *there is*. Contrary to Luzón,[53] the category of 'Assuming shared experiences/knowledge, goals, beliefs' encompasses instances of modality (including *can*, in phrases such as *we can find*). This includes deontic modals or verbs of volition, etc. when they are *not* used to express a statement of opinion. In this latter case, instances are classified as 'Assuming shared experiences/knowledge, goals, beliefs'. Examples are primarily those where *wish*/*want* occur in hypothetical or non-actual contexts, such as *if we want to*, *when we wish to*, *whenever we want*, and *that we want*. In such cases, the writer draws on assumed shared wishes or desires. In other cases, deontic modals are classified as 'Expressing opinion or volition', as explained below. To put simply, the deciding aspect for categorising instances is always the rhetorical function of *we* and its co-text.

In addition, the first-person plural pronoun *we* is also used to **draw or call the reader's attention** to a specific aspect, such as in *we note that* (Luzón, 2009: p. 202). Other common verbs used for this purpose include *emphasize*, *notice*, *point out*, and *stress*. This fifth function is somewhat similar to guiding the reader through the text, except that the use of specific verbs (i.e. *emphasize*, *notice* etc.) is mostly seen to co-occur with adjectives such as *important*, *interesting*, or with the modal verb *should*, in capturing reader's attention.

[53] Luzón (2009) includes modality within two categories: Assuming shared experiences/knowledge, goals, beliefs; and Expressing volition. In the former sub-category, she enjoins co-occurrences of *we* with *need to*, *have to*, *don't need to*, *want to*, *should* etc.; whereas I will enlist them under the latter category to avoid confusion.

The last discourse function of *we* identified by Luzón (2009) is **expressing opinion or volition**. This is where I adapt her scheme slightly, by including the co-occurrence of the first-person plural pronoun with verbs expressing necessity (*need*), opinion (*think*, *believe*), volition (*want*, *wish*), and deontic modality (*should*, *have to*). More specifically, use of *we* in this manner indicates the writer's intention to share an opinion, view or attitude (for example by expressing agreement, disagreement or interest) with regard to known information or established facts (Tang & John, 1999: p. 28).

5.3.2 Functions of *we* in MCSAW

To investigate the many types of *we* identified in the Malaysian corpus, 10% of the total *we* occurrences (3,148) were randomly selected using WordSmith Tool's function 'random thinning' (similar to the qualitative analysis of *can* in Chapter 4). In so doing, each concordance line for the 315 *we* instances is analysed. Table 5.7 shows the discourse functions of *we* in MCSAW. As can be seen, *we* is used most in assuming shared experiences/knowledge, goals, beliefs (86.3%). This is followed by expression of opinion or volition (8.9%). By contrast, Malaysian learners appear to use *we* less frequently in stating conclusions (1.6%), calling the reader's attention (1.3%), and in guiding the reader through the text as well as for stating purposes (both 1.0%).

Table 5.7: Discourse function of *we* in MCSAW (%)

Discourse Function	**Raw**	**%**
Stating the thesis	3	1.0
Stating conclusions	5	1.6
Guiding the reader through the text	3	1.0
Assuming shared experiences/knowledge, goals, beliefs	272	86.3
Emphasising or calling the reader's attention	4	1.3
Expressing opinion or volition	28	8.9
Total	315	100

5.3.2.1 Stating the thesis

As seen in Chapter 3, the core component of an argumentative essay is firstly described in the development of a position, which is also regarded as the development of an argument (Wingate, 2012). This can be identified as the statement of thesis. During this stage, the discourse topic is introduced and the writer's proposition or central statement is made (Hyland, 1990). Hyland further mentions that the proposition is central in the thesis stage, as it "functions to furnish a specific statement of position which defines the topic and gives a focus to the entire composition" (ibid: pp. 70-71). However, closer inspection of the texts in the present study revealed that it is almost too difficult to identify the proper thesis statement in MCSAW essays, and therefore the proposition is often unclear.

Furthermore, most of the essays in the present study have very short introductions that do not state the purpose nor declare the writer's position of the argument clearly. There are only three examples among the 315 concordances that reflect the use of *we* in stating the topic, and they are shown to occur in the introduction of the essay; at the end of the introductory paragraph. The topic is also identifiable due to the use of prompt words such as *advantages* and *disadvantages*, which explicitly signals the essay topics. In the following example (1), the writer states that there are two benefits of using Facebook, whereas in (2) and (3), the writer proposes that there are both advantages and disadvantages of using Facebook:

(1) Nowadays, one of social networking website known as Facebook are knowledgeable among all people around the world. Especially, the teenagers. Facebook have many function and worth when using it in a right ways. I agree with this statement. Actually, Facebook have many advantages than disadvantages. *There are two advantages <u>we</u> can get from using this social networking website known as Facebook* (408.txt).

(2) Social networking has been a common use on the internet in this generation. Facebook has over millions of members to connecting with their friends every day. However, Facebook has been become who was the member's daily life. *<u>We</u> have advantages and disadvantages of using Facebook* (1046.txt).

(3) Nowadays, there are many social network can be found using the internet and many people are using it. With internet, worldwide can use it to find information or using the internet to fulfill their needed. Social network have made people attracted to use the internet without limit. One of the most popular social networking is facebook. The popularity of facebook was increased drastically among people all over the world. In

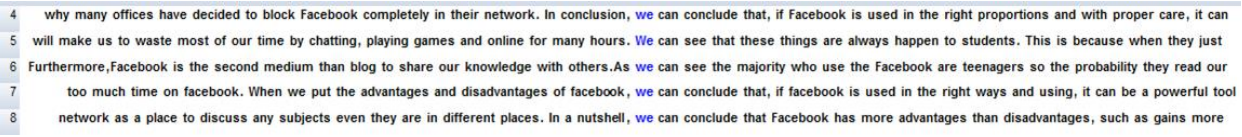
> the other hand, facebook has become very important part of our life. Sometime it can help us in many ways and sometime it can harm us. *There are many advantages and disadvantages of facebook that* <u>*we*</u> *must know for our betterment* (186.txt).

The instances of *we* in these examples can be rephrased as *There are* (*two*) *advantages* (*and disadvantages*) *of using Facebook*. Drawing on Martin and Rose (2003), such topic sentences work as a 'macro-Theme', which functions to construct the development of a text, usually hinting at the 'periodicity', i.e. rhythm of discourse or information flow. Periodicity is regarded as "the layers of prediction that flag for readers what's to come, and the layers of consolidation that accumulate the meanings made" (Martin & Rose, 2003: p. 17). The strategy of predicting phases of discourse with 'macro-Themes', therefore, is a way in which these writers organise their texts in order for readers to process meanings from the texts.

It can also be seen that *we* is used with verbs such as *get*, *have* and *know*. Biber et al. (2002) note that the activity verb *get* and mental/cognitive verb *know* are listed among the twelve verbs that are most common in English. They further add that the transitive main verb *have* is "as common as the most common lexical verbs in English" (ibid: p. 136). For example, the writer of (2) has chosen *we have advantages and disadvantages of using Facebook* instead of using an existential sentence construction such as *There are* (*two*) *advantages* (*and disadvantages*) *of using Facebook* to introduce the thesis. According to Lee and Chen (2009: p. 154), these are not specific to academic writing but of more general currency. In fact, these three types of verbs are considered to be common in conversation (Biber et al., 2002). One explanation could be that learners are not taught explicitly how to produce statements of thesis in the classroom, or that they have limited vocabulary to express such statements. This would also explain the difficulty in identifying thesis statements in the learner corpus in general. As regards the Malay language, it is common for writers to use existential statements in formal writing such as essays written in the classroom. However, it could be possible that learners are influenced by speech in their writing, since the personal plural pronoun *we* is found in all three thesis statements above.

5.3.2.2 Stating conclusions

There are five expressions (1.6%) with *we* that signal conclusions in the 315 instances in MCSAW, as shown in Figure 5.4, including three occurrences of *we can conclude*. As explained in Hyland (1990), the conclusion stage "functions to consolidate the discourse and retrospectively affirm what has been communicated" (p. 74). This usually includes transition signals (e.g. *thus*, *therefore, to conclude*), a restatement of the themes of arguments to the proposition (usually an affirmation of the proposition), and/or a prospective focus.



```
4     why many offices have decided to block Facebook completely in their network. In conclusion, we can conclude that, if Facebook is used in the right proportions and with proper care, it can
5   will make us to waste most of our time by chatting, playing games and online for many hours. We can see that these things are always happen to students. This is because when they just
6 Furthermore,Facebook is the second medium than blog to share our knowledge with others.As we can see the majority who use the Facebook are teenagers so the probability they read our
7          too much time on facebook. When we put the advantages and disadvantages of facebook, we can conclude that, if facebook is used in the right ways and using, it can be a powerful tool
8        network as a place to discuss any subjects even they are in different places. In a nutshell, we can conclude that Facebook has more advantages than disadvantages, such as gains more
```

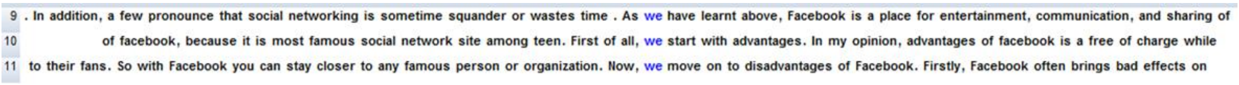
Figure 5.4: Concordance lines for *we* stating conclusions

One observation is the phrase *we can conclude* occurring after the transition markers, *In conclusion* and *In a nutshell* in concordance lines (4) and (8), respectively. This brings about a sense of tautology or redundancy, because the use of *we* in the phrase *we can conclude* repeats the same meaning as the transition markers. The phrase *we can conclude that* is also found to occur after a subordinate/complement clause (in line 7). It is found that the inclusive *we* is used in both the subordinate clause *When <u>we</u> put the advantages and disadvantages of Facebook* and in the continuing phrase *<u>we</u> can conclude that*. In spite of the inclusive pronouns used to construct solidarity between writer and reader, the occurrence of *we* after the conditional *when* and the recurrence of *we* in the following clause raises greater writer/reader visibility in writing.

In addition, two instances show a use of *we* that provides interim conclusions within the essay (lines 5 and 6). In line 5, the active phrase *we can see that* could be replaced with the passive construction *it can be seen that* – concluding or summarising the effects of Facebook on students. On another note, Biber et al. (2002: p. 315) state that the co-occurrence of the mental/cognitive verb *see* with a *that*-clause is more common in fiction (over 200 per million words) compared to academic prose (over 100 per million words). Line 6, on the other hand, shows the use of *we* in the phrase *as we can see*, to realise a sense of commonality between writer and reader for the statement *majority who use the Facebook are teenagers*, in bringing a close to the whole argument for the essay. As noted by Lee and Chen (2009), combinations of *we can*, *we can see*, and *we can see that* are significantly more frequent in learner writing.

5.3.2.3 Guiding the reader through the text

One of the least frequent uses of *we* in the learner corpus (1%) is in statements that guide the reader through the text. As mentioned in Luzón (2009), these statements indicate and direct readers through the structure of the text. Hyland (1990: p. 72) explains that this is among the features in the argument stage, in which "[t]he marker frames the sequence and connects it to both the steps in the argument and to the proposition", often to explicitly guide the reader through the argument stage. He adds that there are two ways of realising this: usually by listing signals such as 'first(ly)', 'second(ly)', 'next', etc.; and transition signals that indicate the step to another sequence (i.e. adverbial connectives, conjunctions and comments indicating changes in the discussion). Figure 5.5 presents the three instances of *we* functioning to guide readers in MCSAW texts.



```
 9  . In addition, a few pronounce that social networking is sometime squander or wastes time . As we have learnt above, Facebook is a place for entertainment, communication, and sharing of
10                of facebook, because it is most famous social network site among teen. First of all, we start with advantages. In my opinion, advantages of facebook is a free of charge while
11   to their fans. So with Facebook you can stay closer to any famous person or organization. Now, we move on to disadvantages of Facebook. Firstly, Facebook often brings bad effects on
```

Figure 5.5: Concordance lines for *we* as guiding reader through text

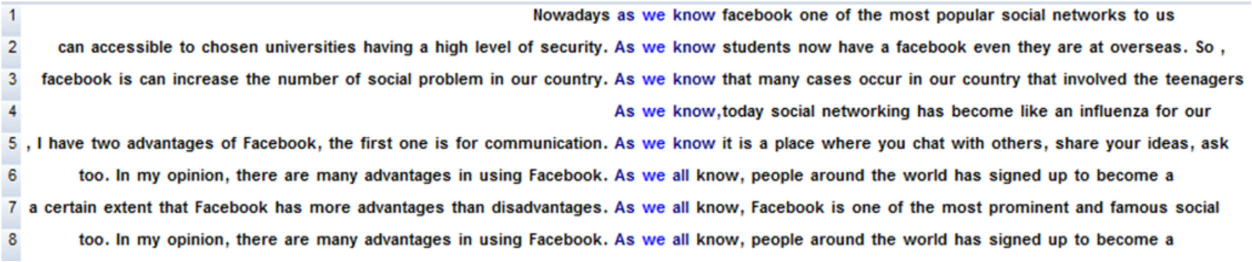
The first line (line 9) refers to the previous context by use of the phrase *as <u>we</u> have learnt above*, and therefore is used anaphorically. It is also found that the pronoun *we* occurs in the proximity of another 'signpost', *first of all*, in line 10. The use of *we* in *<u>First of all</u>, <u>we</u> start with advantages* signals the writer's intention to begin his or her argument with the advantages of Facebook. Line 11, on the other hand, indicates a change of focus from discussing advantages of Facebook to the disadvantages of Facebook. This is complemented with the transition signal *Now* and activity verb *move on*, that indicate changes in the discussion, *<u>Now</u>, <u>we</u> move on to disadvantages of Facebook*. As stated in Gilquin and Paquot (2007: pp. 4-5) and Granger and Rayson (1998: p. 124), spoken-like lexical items such as 'first of all' and 'now' are over-used by learners. These, according to Gilquin and Paquot (2007), are described as emphasisers, and mark learners' writing as spoken-like, since emphasisers are more common in speech. Nevertheless, findings in MCSAW reveal that learners under-use the personal pronoun *we* for this particular discourse function, and when employed, their writing tends to resemble speech written down.

5.3.2.4 Assuming shared experiences/knowledge, goals, beliefs

The majority use of *we* in MCSAW is made up of this function, of 'Assuming shared experiences/knowledge, goals, beliefs' (86.3%). More specifically, it can be found that *we* in most of these cases is used as a generic first-person pronoun or as a substitute representing a larger group of people (Tang & John, 1999). As mentioned above, instances that fall within this category basically encompass assumptions of shared experience and some instances of modality (*we can find*), including cases where deontic modals or verbs of volition, etc. are not used to express a statement of opinion. Examples are primarily those where *wish*/*want* occur in hypothetical or non-actual contexts, such as *if we want to*; *when we wish to*; *whenever we want*; *that we want*.

Given the most number of instances occurring in this category, sub-categories are made to further group them. It is found that 8 instances incorporate the phrases *as we know* and *as we all know*. 45 instances are grouped under hypothetical or non-actual contexts. There are 152 instances that denote 'we can' phrases; while the remaining 67 instances express either advantages, disadvantages, or neutral meanings with regard to the essay topics. Where instances are more than 20, only 20 concordance lines will be shown for ease of reading.

As regards the first 8 instances of *we* used in assuming shared experiences, writers use the phrases *as we know* and *as we all know*, as mentioned and found in Hyland (2002a) and Fawcett (2013). Instances such as 'as we know' or 'as we all know' incorporate the inclusive *we* and indicate the assumption of readers' agreement on or shared experiences of a proposition (Luzón, 2009). Figure 5.6 shows instances of *we* in the two common phrases that indicate the intention to involve the reader, and hence creating the assumption of shared experiences. Note that lines 6 and 8 are similar to each other, and may suggest plagiarism on the part of the students or effects of classroom teaching, since these texts are extracted from separate files.

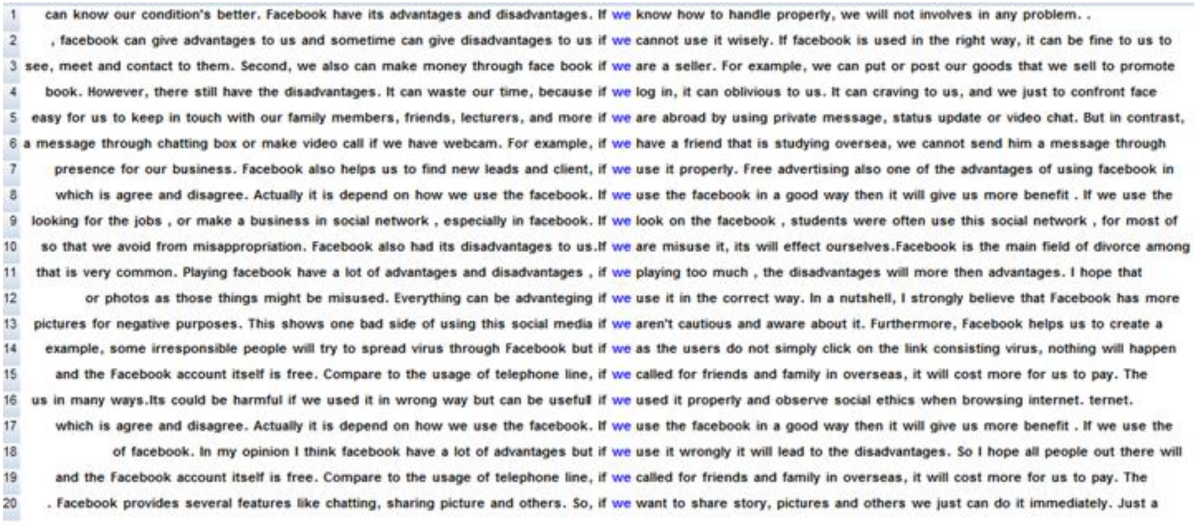


```
1                                                         Nowadays as we know facebook one of the most popular social networks to us
2      can accessible to chosen universities having a high level of security. As we know students now have a facebook even they are at overseas. So ,
3   facebook is can increase the number of social problem in our country. As we know that many cases occur in our country that involved the teenagers
4                                                                    As we know,today social networking has become like an influenza for our
5 , I have two advantages of Facebook, the first one is for communication. As we know it is a place where you chat with others, share your ideas, ask
6         too. In my opinion, there are many advantages in using Facebook. As we all know, people around the world has signed up to become a
7  a certain extent that Facebook has more advantages than disadvantages. As we all know, Facebook is one of the most prominent and famous social
8         too. In my opinion, there are many advantages in using Facebook. As we all know, people around the world has signed up to become a
```

Figure 5.6: Concordance lines for *as we know* and *as we all know*

As mentioned earlier, Fawcett (2013) also found such uses among Chinese learners' argumentative essays, and argued that learners tend to use these phrases in making claims involving a general consensus, which express learners' inherent need "to qualify their assertions with extensive personal reference, mitigating the strength and authority of their statements" (Fawcett, 2013: p. 268). Use of phrases *as we know*/*as we all know*, in turn, illustrates learners' attempts at constructing personalised forms, despite being able to express certainty, that suggests their lack of control of this genre (Hyland & Milton, 1997). In many cases, Luzón (2009: p. 201) argues, the "passive voice or sentences with inanimate subjects would be preferred in expert writing", and therefore *it has been known*/*it is known that* would be favoured in the above instances.

It is also found that there are 45 instances of *we* used to show shared experiences/knowledge in hypothetical or non-actual contexts. These include instances with conditional constructions such as 'if we', 'when…we', and 'when/whenever we'. Figure 5.7 presents 20 instances of the 'if we' construction, generally functioning to assume shared experience/knowledge between the writer and reader.

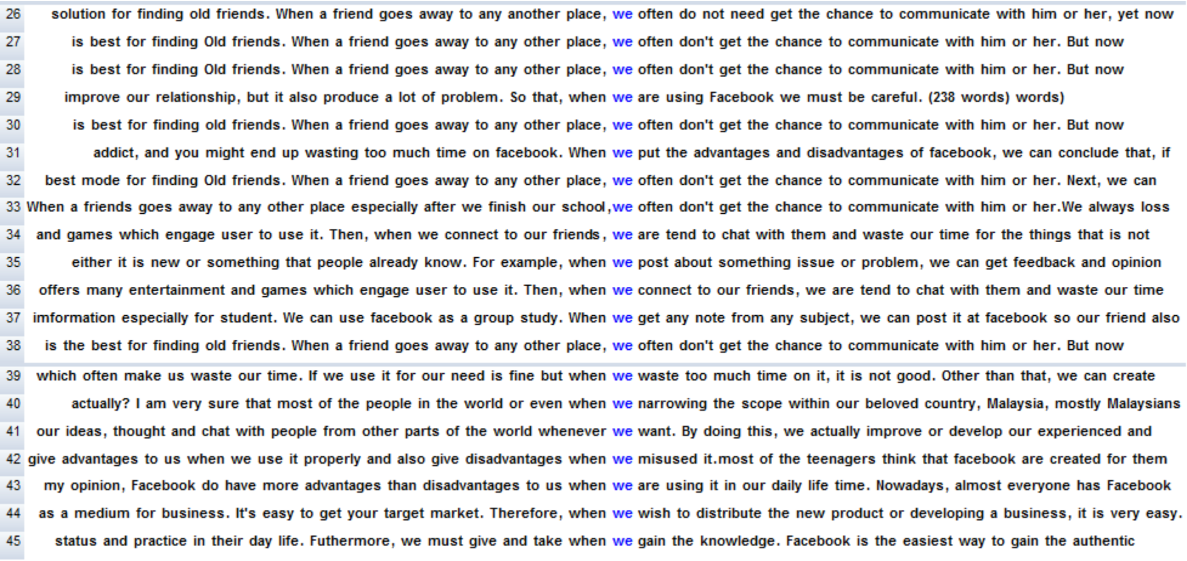


```
1     can know our condition's better. Facebook have its advantages and disadvantages. If we know how to handle properly, we will not involves in any problem. .
2      , facebook can give advantages to us and sometime can give disadvantages to us if we cannot use it wisely. If facebook is used in the right way, it can be fine to us to
3  see, meet and contact to them. Second, we also can make money through face book if we are a seller. For example, we can put or post our goods that we sell to promote
4     book. However, there still have the disadvantages. It can waste our time, because if we log in, it can oblivious to us. It can craving to us, and we just to confront face
5   easy for us to keep in touch with our family members, friends, lecturers, and more if we are abroad by using private message, status update or video chat. But in contrast,
6 a message through chatting box or make video call if we have webcam. For example, if we have a friend that is studying oversea, we cannot send him a message through
7      presence for our business. Facebook also helps us to find new leads and client, if we use it properly. Free advertising also one of the advantages of using facebook in
8      which is agree and disagree. Actually it is depend on how we use the facebook. If we use the facebook in a good way then it will give us more benefit . If we use the
9   looking for the jobs , or make a business in social network , especially in facebook. If we look on the facebook , students were often use this social network , for most of
10   so that we avoid from misappropriation. Facebook also had its disadvantages to us.If we are misuse it, its will effect ourselves.Facebook is the main field of divorce among
11  that is very common. Playing facebook have a lot of advantages and disadvantages , if we playing too much , the disadvantages will more then advantages. I hope that
12           or photos as those things might be misused. Everything can be advanteging if we use it in the correct way. In a nutshell, I strongly believe that Facebook has more
13  pictures for negative purposes. This shows one bad side of using this social media if we aren't cautious and aware about it. Furthermore, Facebook helps us to create a
14    example, some irresponsible people will try to spread virus through Facebook but if we as the users do not simply click on the link consisting virus, nothing will happen
15     and the Facebook account itself is free. Compare to the usage of telephone line, if we called for friends and family in overseas, it will cost more for us to pay. The
16  us in many ways.Its could be harmful if we used it in wrong way but can be usefull if we used it properly and observe social ethics when browsing internet. ternet.
17     which is agree and disagree. Actually it is depend on how we use the facebook. If we use the facebook in a good way then it will give us more benefit . If we use the
18              of facebook. In my opinion I think facebook have a lot of advantages but if we use it wrongly it will lead to the disadvantages. So I hope all people out there will
19     and the Facebook account itself is free. Compare to the usage of telephone line, if we called for friends and family in overseas, it will cost more for us to pay. The
20    . Facebook provides several features like chatting, sharing picture and others. So, if we want to share story, pictures and others we just can do it immediately. Just a
```

Figure 5.7: Concordance lines for *if we* constructions

These instances refer to the use of conjunction *if* in marking conditions in the arguments put forth by learners, usually followed by a consequential clause. This, in turn, increases learners' tendency to re-use the personal plural pronoun (which also explains for the overuse of *we*), instead of using impersonal structures with the main subject at the beginning of the sentence. Such examples include instances that could be rephrased without having to employ the inclusive *we*. For instance, the example in line 2 could be rephrased as *Facebook provides advantages and at times disadvantages if users do not use it wisely*, while example 15 could be rephrased as *It would be costly to call friends and family overseas*. Furthermore, Biber et al. (2002) note that *if*-clauses of condition are particularly more frequent in conversations.

Figure 5.8 shows 20 examples of *we* co-occurring with conditional *when*, in two types: lines 26-30, 32-34, 38 and 45 demonstrate that, when X happens, we do (or not do) Y (e.g. *when a friend goes away, we don't get the chance...*); whereas lines 35-37, 39-40, and 42-44 exemplify that, when we do X, we receive some form of response (e.g. *when we post…we can get feedback*). In both situations, "[a] hypothetical condition implies that the condition is not fulfilled" (Biber et al., 2002: p. 374). In addition, it can be seen that a recurring sentence is used in lines 27, 28, 30, 32, 33, and 38 (*when a friend goes away to any other place, we often don't get the chance to communicate with him or her*). This is another example of potential plagiarism on the students' part, or a prompt sentence used in the classroom. Figure 5.8 also shows *we* co-occurring immediately to the right position of the conditional *whenever*. This also describes the non-actual context, where line 41 conveys the experience of interacting with people around the world being achievable at any time.



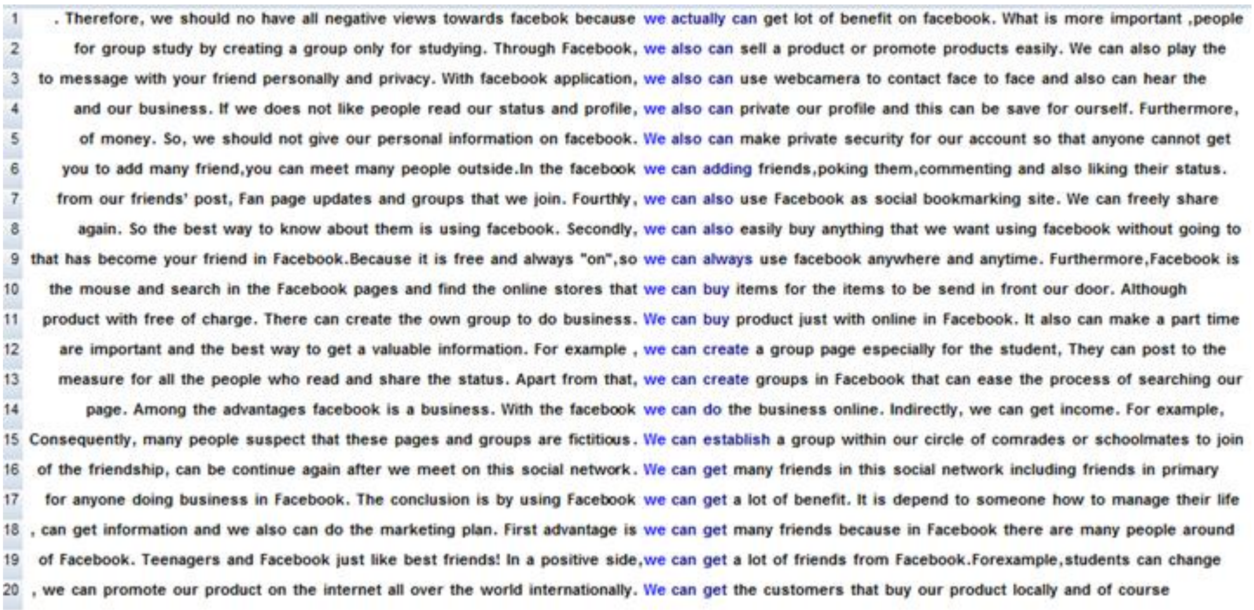
```
26      solution for finding old friends. When a friend goes away to any another place, we often do not need get the chance to communicate with him or her, yet now
27         is best for finding Old friends. When a friend goes away to any other place, we often don't get the chance to communicate with him or her. But now
28         is best for finding Old friends. When a friend goes away to any other place, we often don't get the chance to communicate with him or her. But now
29       improve our relationship, but it also produce a lot of problem. So that, when we are using Facebook we must be careful. (238 words) words)
30         is best for finding old friends. When a friend goes away to any other place, we often don't get the chance to communicate with him or her. But now
31            addict, and you might end up wasting too much time on facebook. When we put the advantages and disadvantages of facebook, we can conclude that, if
32     best mode for finding Old friends. When a friend goes away to any other place, we often don't get the chance to communicate with him or her. Next, we can
33 When a friends goes away to any other place especially after we finish our school,we often don't get the chance to communicate with him or her.We always loss
34   and games which engage user to use it. Then, when we connect to our friends, we are tend to chat with them and waste our time for the things that is not
35         either it is new or something that people already know. For example, when we post about something issue or problem, we can get feedback and opinion
36    offers many entertainment and games which engage user to use it. Then, when we connect to our friends, we are tend to chat with them and waste our time
37  imformation especially for student. We can use facebook as a group study. When we get any note from any subject, we can post it at facebook so our friend also
38     is the best for finding old friends. When a friend goes away to any other place, we often don't get the chance to communicate with him or her. But now
39   which often make us waste our time. If we use it for our need is fine but when we waste too much time on it, it is not good. Other than that, we can create
40         actually? I am very sure that most of the people in the world or even when we narrowing the scope within our beloved country, Malaysia, mostly Malaysians
41   our ideas, thought and chat with people from other parts of the world whenever we want. By doing this, we actually improve or develop our experienced and
42 give advantages to us when we use it properly and also give disadvantages when we misused it.most of the teenagers think that facebook are created for them
43    my opinion, Facebook do have more advantages than disadvantages to us when we are using it in our daily life time. Nowadays, almost everyone has Facebook
44   as a medium for business. It's easy to get your target market. Therefore, when we wish to distribute the new product or developing a business, it is very easy.
45      status and practice in their day life. Futhermore, we must give and take when we gain the knowledge. Facebook is the easiest way to gain the authentic
```

Figure 5.8: Concordance lines for *when/whenever…we* constructions

Essentially, the three types of condition clauses serve special conversational uses, especially in giving suggestions (e.g. *if we know how to handle properly, we will not involves in any problem*; *when we are using Facebook, we must be careful*; *If we use it for our need is fine but when we waste too much time on it, it is not good*). According to Biber et al. (2002: p. 375), "the use of a conditional clause can soften the suggestion or command", and in turn, learners may have used these clauses in their argument to suggest, or to persuade readers. More specifically, the use of *we* functioning to assume shared experiences/knowledge in the hypothetical sense, as depicted in Figure 5.7 and Figure 5.8, demonstrates the use of condition clauses that include subordinators expressing meanings such as time (*when*, *whenever*) and condition (*if*). Meanwhile, Biber et al. (2002: pp. 375-376) assert that "registers show interesting preferences for certain semantic categories", and that, while condition clauses are highly frequent in conversation, purpose clauses, i.e. <u>*In order to help such children*</u>, *it is necessary to introduce novel and artificial procedures to assist learning* (Biber et al.'s example) are notably more common in academic prose, where they help to explain recommendations.

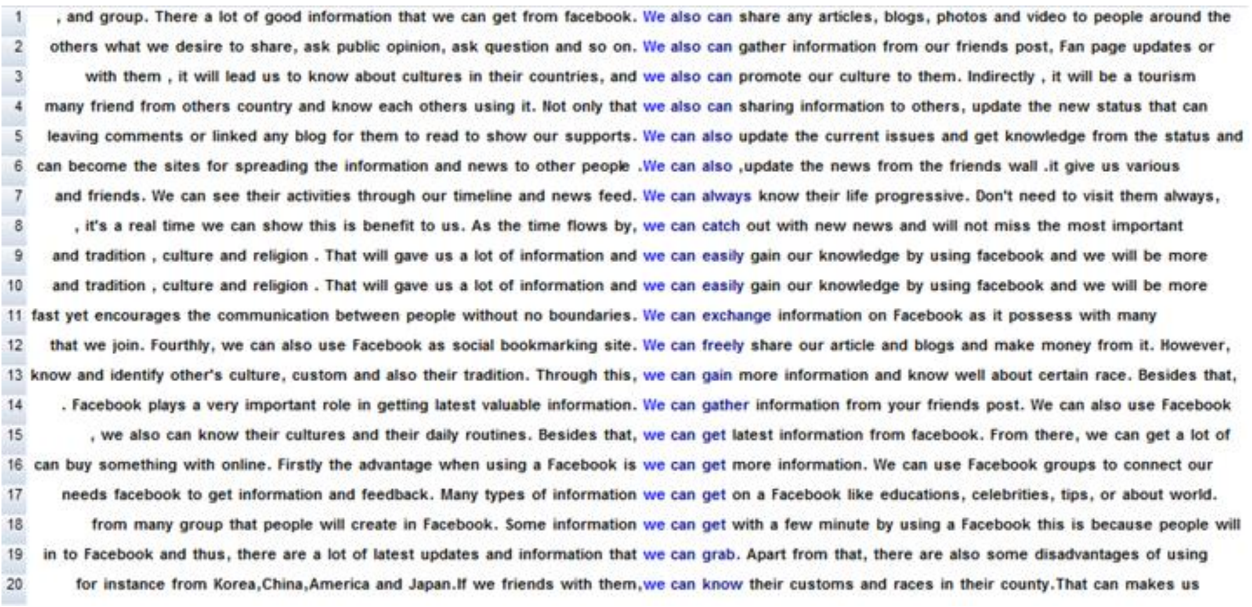
Moving on, the majority of instances of *we* that indicate an assumption of shared experiences/knowledge, goals, beliefs (145 occurrences) were found to co-occur with the modal verb *can*. This should not be surprising, as it has been found, in the previous chapter, that both *we* and *can* are highly significant collocates of each other. According to Luzón (2009), *we* is found to be used with modal *can* as "a proxy representing a larger group of people" (p. 201): for example, *This page is orderly and we can understand it quite well* could be replaced with *This page is orderly and easy to understand*. It is also found in sentences such as *Nowadays we can find a wide range of different displays* (Luzón's example), which shows an existential meaning. However, *we* co-occurs more frequently with *can* in MCSAW than in the findings revealed in Luzón (2009). As discussed in Chapter 4, *can* in MCSAW is often used to denote the 'Ability' meaning. Coupled with the personal pronoun *we*, which is almost always inclusive, the *we* + *can* phrase suggests the ability of the collective/group of people to realise a particular action. More specifically, this is seen as a type of persuasion to support the validity of the proposition, that is, "a statement appealing to the potency of 'shared' presuppositions or expectations about topic background, presenting a generalization based on factual evidence or expert opinion, or a declaration of opinion" (Hyland, 1990: pp. 72-73). In turn, the possibility or ability meanings that are derived from these instances are seen to be expressed by personal forms, in which shared knowledge between reader and writer is assumed or expected.

For example, Figure 5.9 presents 20 of 45 *we* + *can* instances that relate to different features of Facebook such as selling/buying/promoting products via the site (lines 2, 8, 10, 11), using web-cameras for interaction (line 3), or 'adding' friends on one's profile (line 6). It can also be seen that the *we* + *can* construction co-occurs frequently with the verb *get* in expressing the physical action of acquiring friends/benefits/customers (lines 16-20). Figure 5.10 shows another 20 of 43 *we* + *can* phrases in expressing information sharing/ gathering. This also refers to the physical meanings of the word forms *share, gather*, *promote*, *update,* and *know*. Figure 5.11 shows 20 out of 57 instances of the phrase indicating some sense of connection with others. This is notably seen in words such as *find*, *feedback, search*, *ask, chat, communicate,* and *connect*. Essentially, these instances include 'we can' phrases co-occurring with common simple lexical verbs (e.g. *know*, *keep*, *make*, *get* etc.) and other simple generic verbs that are topic-related, i.e. *share*, *search*, *save*, *post*, *chat*, *communicate*, and *connect*. These phraseological structures indicate that not only is modality frequently in association with the use of the personal plural pronoun *we*, but that the *we* + *can* construction typically is related to the topic of Facebook.



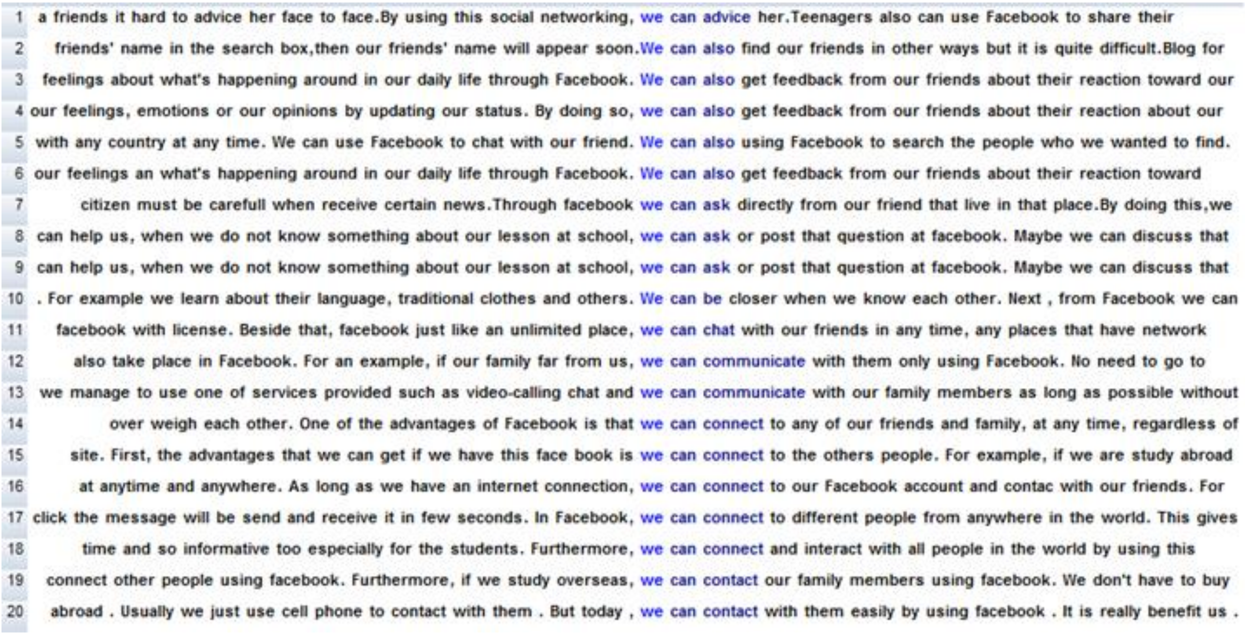
```
1   . Therefore, we should no have all negative views towards facebok because we actually can get lot of benefit on facebook. What is more important ,people
2   for group study by creating a group only for studying. Through Facebook, we also can sell a product or promote products easily. We can also play the
3   to message with your friend personally and privacy. With facebook application, we also can use webcamera to contact face to face and also can hear the
4   and our business. If we does not like people read our status and profile, we also can private our profile and this can be save for ourself. Furthermore,
5   of money. So, we should not give our personal information on facebook. We also can make private security for our account so that anyone cannot get
6   you to add many friend,you can meet many people outside.In the facebook we can adding friends,poking them,commenting and also liking their status.
7   from our friends' post, Fan page updates and groups that we join. Fourthly, we can also use Facebook as social bookmarking site. We can freely share
8   again. So the best way to know about them is using facebook. Secondly, we can also easily buy anything that we want using facebook without going to
9   that has become your friend in Facebook.Because it is free and always "on",so we can always use facebook anywhere and anytime. Furthermore,Facebook is
10  the mouse and search in the Facebook pages and find the online stores that we can buy items for the items to be send in front our door. Although
11  product with free of charge. There can create the own group to do business. We can buy product just with online in Facebook. It also can make a part time
12  are important and the best way to get a valuable information. For example , we can create a group page especially for the student, They can post to the
13  measure for all the people who read and share the status. Apart from that, we can create groups in Facebook that can ease the process of searching our
14  page. Among the advantages facebook is a business. With the facebook we can do the business online. Indirectly, we can get income. For example,
15  Consequently, many people suspect that these pages and groups are fictitious. We can establish a group within our circle of comrades or schoolmates to join
16  of the friendship, can be continue again after we meet on this social network. We can get many friends in this social network including friends in primary
17  for anyone doing business in Facebook. The conclusion is by using Facebook we can get a lot of benefit. It is depend to someone how to manage their life
18  , can get information and we also can do the marketing plan. First advantage is we can get many friends because in Facebook there are many people around
19  of Facebook. Teenagers and Facebook just like best friends! In a positive side,we can get a lot of friends from Facebook.Forexample,students can change
20  , we can promote our product on the internet all over the world internationally. We can get the customers that buy our product locally and of course
```

Figure 5.9: Concordance lines for *we can* relating to different aspects of Facebook



```
1   , and group. There a lot of good information that we can get from facebook. We also can share any articles, blogs, photos and video to people around the
2   others what we desire to share, ask public opinion, ask question and so on. We also can gather information from our friends post, Fan page updates or
3   with them , it will lead us to know about cultures in their countries, and we also can promote our culture to them. Indirectly , it will be a tourism
4   many friend from others country and know each others using it. Not only that we also can sharing information to others, update the new status that can
5   leaving comments or linked any blog for them to read to show our supports. We can also update the current issues and get knowledge from the status and
6   can become the sites for spreading the information and news to other people .We can also ,update the news from the friends wall .it give us various
7   and friends. We can see their activities through our timeline and news feed. We can always know their life progressive. Don't need to visit them always,
8   , it's a real time we can show this is benefit to us. As the time flows by, we can catch out with new news and will not miss the most important
9   and tradition , culture and religion . That will gave us a lot of information and we can easily gain our knowledge by using facebook and we will be more
10  and tradition , culture and religion . That will gave us a lot of information and we can easily gain our knowledge by using facebook and we will be more
11  fast yet encourages the communication between people without no boundaries. We can exchange information on Facebook as it possess with many
12  that we join. Fourthly, we can also use Facebook as social bookmarking site. We can freely share our article and blogs and make money from it. However,
13  know and identify other's culture, custom and also their tradition. Through this, we can gain more information and know well about certain race. Besides that,
14  . Facebook plays a very important role in getting latest valuable information. We can gather information from your friends post. We can also use Facebook
15  , we also can know their cultures and their daily routines. Besides that, we can get latest information from facebook. From there, we can get a lot of
16  can buy something with online. Firstly the advantage when using a Facebook is we can get more information. We can use Facebook groups to connect our
17  needs facebook to get information and feedback. Many types of information we can get on a Facebook like educations, celebrities, tips, or about world.
18  from many group that people will create in Facebook. Some information we can get with a few minute by using a Facebook this is because people will
19  in to Facebook and thus, there are a lot of latest updates and information that we can grab. Apart from that, there are also some disadvantages of using
20  for instance from Korea,China,America and Japan.If we friends with them,we can know their customs and races in their county.That can makes us
```

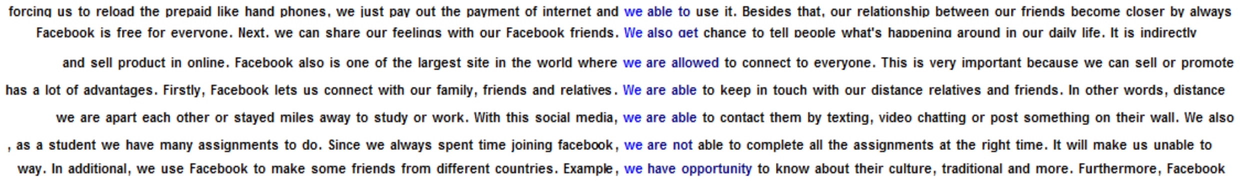
Figure 5.10: Concordance lines for *we can* expressing information sharing/gathering



```
1  a friends it hard to advice her face to face.By using this social networking, we can advice her.Teenagers also can use Facebook to share their
2  friends' name in the search box,then our friends' name will appear soon.We can also find our friends in other ways but it is quite difficult.Blog for
3  feelings about what's happening around in our daily life through Facebook. We can also get feedback from our friends about their reaction toward our
4  our feelings, emotions or our opinions by updating our status. By doing so, we can also get feedback from our friends about their reaction about our
5  with any country at any time. We can use Facebook to chat with our friend. We can also using Facebook to search the people who we wanted to find.
6  our feelings an what's happening around in our daily life through Facebook. We can also get feedback from our friends about their reaction toward
7  citizen must be carefull when receive certain news.Through facebook we can ask directly from our friend that live in that place.By doing this,we
8  can help us, when we do not know something about our lesson at school, we can ask or post that question at facebook. Maybe we can discuss that
9  can help us, when we do not know something about our lesson at school, we can ask or post that question at facebook. Maybe we can discuss that
10 . For example we learn about their language, traditional clothes and others. We can be closer when we know each other. Next , from Facebook we can
11 facebook with license. Beside that, facebook just like an unlimited place, we can chat with our friends in any time, any places that have network
12 also take place in Facebook. For an example, if our family far from us, we can communicate with them only using Facebook. No need to go to
13 we manage to use one of services provided such as video-calling chat and we can communicate with our family members as long as possible without
14 over weigh each other. One of the advantages of Facebook is that we can connect to any of our friends and family, at any time, regardless of
15 site. First, the advantages that we can get if we have this face book is we can connect to the others people. For example, if we are study abroad
16 at anytime and anywhere. As long as we have an internet connection, we can connect to our Facebook account and contac with our friends. For
17 click the message will be send and receive it in few seconds. In Facebook, we can connect to different people from anywhere in the world. This gives
18 time and so informative too especially for the students. Furthermore, we can connect and interact with all people in the world by using this
19 connect other people using facebook. Furthermore, if we study overseas, we can contact our family members using facebook. We don't have to buy
20 abroad . Usually we just use cell phone to contact with them . But today , we can contact with them easily by using facebook . It is really benefit us .
```

Figure 5.11: Concordance lines for *we can* expressing connection with others

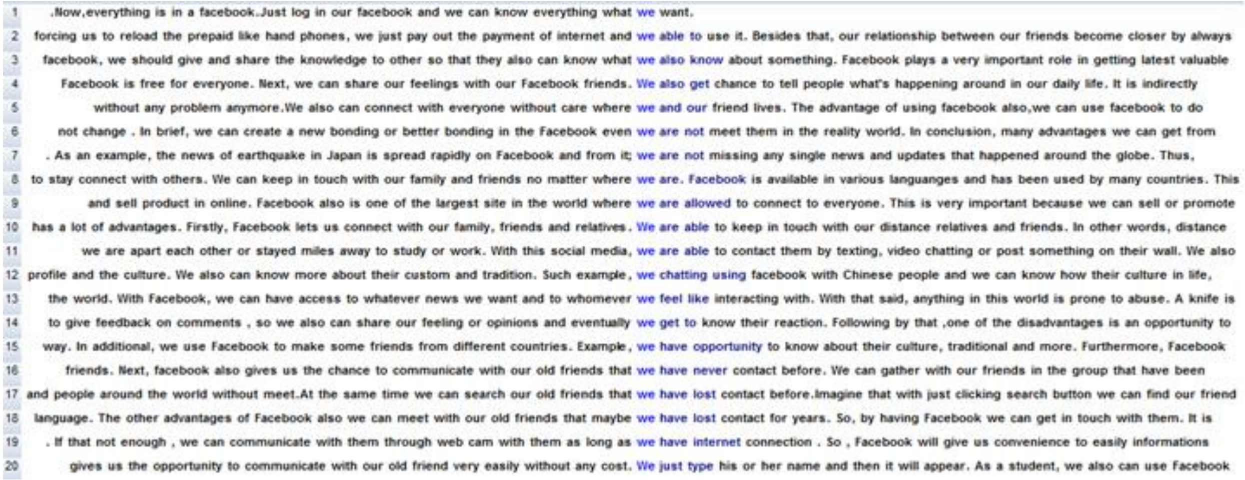
In addition to instances that include *we can*, there are seven occurrences where expressions with similar meanings are used, such as the phrases *we able to*, *we also get chance to*, *we are allowed to, we are able to*, *we have opportunity to* (Figure 5.12). Contrary to the highly frequent 'we can', these alternative constructions are under-used.



```
forcing us to reload the prepaid like hand phones, we just pay out the payment of internet and we able to use it. Besides that, our relationship between our friends become closer by always
Facebook is free for everyone. Next, we can share our feelings with our Facebook friends. We also get chance to tell people what's happening around in our daily life. It is indirectly
and sell product in online. Facebook also is one of the largest site in the world where we are allowed to connect to everyone. This is very important because we can sell or promote
has a lot of advantages. Firstly, Facebook lets us connect with our family, friends and relatives. We are able to keep in touch with our distance relatives and friends. In other words, distance
we are apart each other or stayed miles away to study or work. With this social media, we are able to contact them by texting, video chatting or post something on their wall. We also
, as a student we have many assignments to do. Since we always spent time joining facebook, we are not able to complete all the assignments at the right time. It will make us unable to
way. In additional, we use Facebook to make some friends from different countries. Example, we have opportunity to know about their culture, traditional and more. Furthermore, Facebook
```

Figure 5.12: Related meanings of *we can*

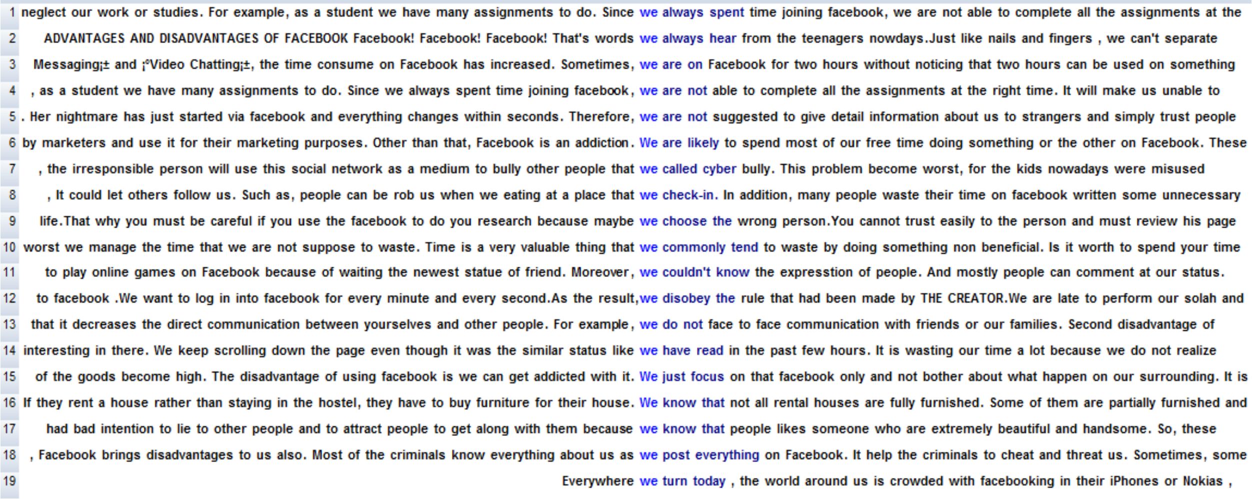
The remaining 67 instances that function to assume shared experience/knowledge can further be sub-categorised, into expressions of advantages/benefits/pros, disadvantages/cons, and neutral expressions in the argumentative essay. This can be seen in instances indicating benefits of Facebook, for example in line 7 of Figure 5.13, …*we are not missing any single news and updates*…, and line 20 of the same Figure, *We just type his or her name and then it will appear*.



```
1        .Now,everything is in a facebook.Just log in our facebook and we can know everything what we want.
2   forcing us to reload the prepaid like hand phones, we just pay out the payment of internet and we able to use it. Besides that, our relationship between our friends become closer by always
3     facebook, we should give and share the knowledge to other so that they also can know what we also know about something. Facebook plays a very important role in getting latest valuable
4         Facebook is free for everyone. Next, we can share our feelings with our Facebook friends. We also get chance to tell people what's happening around in our daily life. It is indirectly
5               without any problem anymore.We also can connect with everyone without care where we and our friend lives. The advantage of using facebook also,we can use facebook to do
6          not change . In brief, we can create a new bonding or better bonding in the Facebook even we are not meet them in the reality world. In conclusion, many advantages we can get from
7      . As an example, the news of earthquake in Japan is spread rapidly on Facebook and from it; we are not missing any single news and updates that happened around the globe. Thus,
8   to stay connect with others. We can keep in touch with our family and friends no matter where we are. Facebook is available in various languanges and has been used by many countries. This
9               and sell product in online. Facebook also is one of the largest site in the world where we are allowed to connect to everyone. This is very important because we can sell or promote
10  has a lot of advantages. Firstly, Facebook lets us connect with our family, friends and relatives. We are able to keep in touch with our distance relatives and friends. In other words, distance
11             we are apart each other or stayed miles away to study or work. With this social media, we are able to contact them by texting, video chatting or post something on their wall. We also
12 profile and the culture. We also can know more about their custom and tradition. Such example, we chatting using facebook with Chinese people and we can know how their culture in life,
13      the world. With Facebook, we can have access to whatever news we want and to whomever we feel like interacting with. With that said, anything in this world is prone to abuse. A knife is
14      to give feedback on comments , so we also can share our feeling or opinions and eventually we get to know their reaction. Following by that ,one of the disadvantages is an opportunity to
15     way. In additional, we use Facebook to make some friends from different countries. Example, we have opportunity to know about their culture, traditional and more. Furthermore, Facebook
16          friends. Next, facebook also gives us the chance to communicate with our old friends that we have never contact before. We can gather with our friends in the group that have been
17 and people around the world without meet.At the same time we can search our old friends that we have lost contact before.Imagine that with just clicking search button we can find our friend
18  language. The other advantages of Facebook also we can meet with our old friends that maybe we have lost contact for years. So, by having Facebook we can get in touch with them. It is
19     . If that not enough , we can communicate with them through web cam with them as long as we have internet connection . So , Facebook will give us convenience to easily informations
20          gives us the opportunity to communicate with our old friend very easily without any cost. We just type his or her name and then it will appear. As a student, we also can use Facebook
```

Figure 5.13: Concordance lines for *we* expressing shared knowledge of advantages

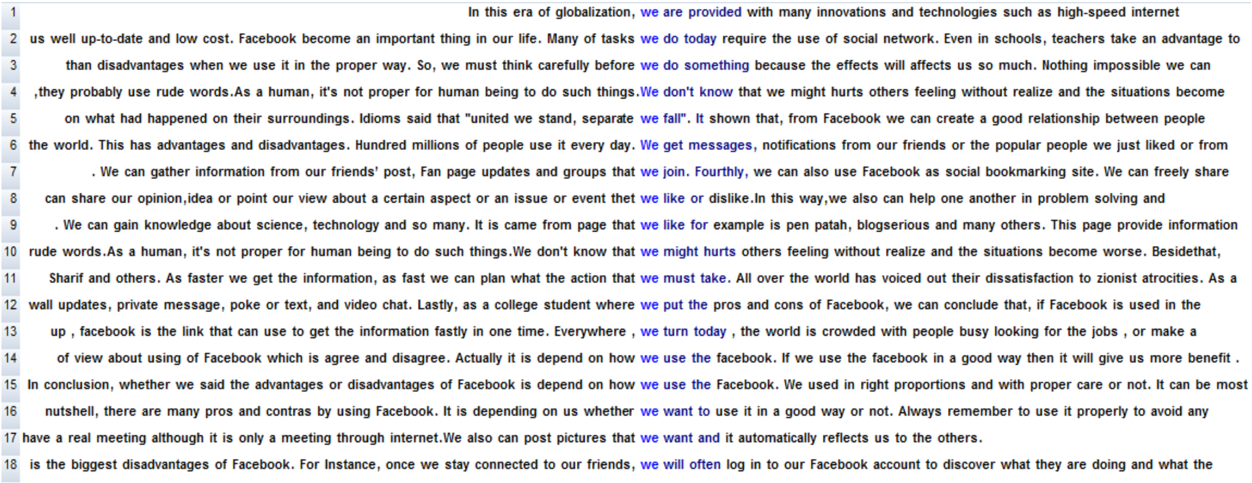
In contrast, Figure 5.14 shows 19 instances that co-textually refer to Facebook as bringing disadvantages to users, such as wasting time (lines 1, 3, 4, 6, 10, 12 and 14) and imposing safety threats (lines 5, 7, 8, 9, 17 and 18).



```
1 neglect our work or studies. For example, as a student we have many assignments to do. Since we always spent time joining facebook, we are not able to complete all the assignments at the
2     ADVANTAGES AND DISADVANTAGES OF FACEBOOK Facebook! Facebook! Facebook! That's words we always hear from the teenagers nowdays.Just like nails and fingers , we can't separate
3   Messaging¡± and ¡°Video Chatting¡±, the time consume on Facebook has increased. Sometimes, we are on Facebook for two hours without noticing that two hours can be used on something
4   , as a student we have many assignments to do. Since we always spent time joining facebook, we are not able to complete all the assignments at the right time. It will make us unable to
5 . Her nightmare has just started via facebook and everything changes within seconds. Therefore, we are not suggested to give detail information about us to strangers and simply trust people
6 by marketers and use it for their marketing purposes. Other than that, Facebook is an addiction. We are likely to spend most of our free time doing something or the other on Facebook. These
7    , the irresponsible person will use this social network as a medium to bully other people that we called cyber bully. This problem become worst, for the kids nowadays were misused
8      , It could let others follow us. Such as, people can be rob us when we eating at a place that we check-in. In addition, many people waste their time on facebook written some unnecessary
9    life.That why you must be careful if you use the facebook to do you research because maybe we choose the wrong person.You cannot trust easily to the person and must review his page
10 worst we manage the time that we are not suppose to waste. Time is a very valuable thing that we commonly tend to waste by doing something non beneficial. Is it worth to spend your time
11     to play online games on Facebook because of waiting the newest statue of friend. Moreover, we couldn't know the expresstion of people. And mostly people can comment at our status.
12   to facebook .We want to log in into facebook for every minute and every second.As the result,we disobey the rule that had been made by THE CREATOR.We are late to perform our solah and
13  that it decreases the direct communication between yourselves and other people. For example, we do not face to face communication with friends or our families. Second disadvantage of
14 interesting in there. We keep scrolling down the page even though it was the similar status like we have read in the past few hours. It is wasting our time a lot because we do not realize
15   of the goods become high. The disadvantage of using facebook is we can get addicted with it. We just focus on that facebook only and not bother about what happen on our surrounding. It is
16 If they rent a house rather than staying in the hostel, they have to buy furniture for their house. We know that not all rental houses are fully furnished. Some of them are partially furnished and
17     had bad intention to lie to other people and to attract people to get along with them because we know that people likes someone who are extremely beautiful and handsome. So, these
18  , Facebook brings disadvantages to us also. Most of the criminals know everything about us as we post everything on Facebook. It help the criminals to cheat and threat us. Sometimes, some
19                                                                                              Everywhere we turn today , the world around us is crowded with facebooking in their iPhones or Nokias ,
```

Figure 5.14: Concordance lines for *we* expressing shared knowledge of disadvantages

18 other instances indicate neutral expressions of Facebook, as shown in Figure 5.15. These are usually generic statements used in the essay to provide context to the topic: for example, *many of tasks <u>we</u> do today require the use of social network* (line 2). Lines 14 and 15 also show instances that indicate shared neutral experiences, in which *we* co-occurs with *depend*, as in *depend on how we use the facebook*.



```
1                                                    In this era of globalization, we are provided with many innovations and technologies such as high-speed internet
2  us well up-to-date and low cost. Facebook become an important thing in our life. Many of tasks we do today require the use of social network. Even in schools, teachers take an advantage to
3      than disadvantages when we use it in the proper way. So, we must think carefully before we do something because the effects will affects us so much. Nothing impossible we can
4  ,they probably use rude words.As a human, it's not proper for human being to do such things.We don't know that we might hurts others feeling without realize and the situations become
5      on what had happened on their surroundings. Idioms said that "united we stand, separate we fall". It shown that, from Facebook we can create a good relationship between people
6  the world. This has advantages and disadvantages. Hundred millions of people use it every day. We get messages, notifications from our friends or the popular people we just liked or from
7           . We can gather information from our friends' post, Fan page updates and groups that we join. Fourthly, we can also use Facebook as social bookmarking site. We can freely share
8    can share our opinion,idea or point our view about a certain aspect or an issue or event thet we like or dislike.In this way,we also can help one another in problem solving and
9     . We can gain knowledge about science, technology and so many. It is came from page that we like for example is pen patah, blogserious and many others. This page provide information
10 rude words.As a human, it's not proper for human being to do such things.We don't know that we might hurts others feeling without realize and the situations become worse. Besidethat,
11    Sharif and others. As faster we get the information, as fast we can plan what the action that we must take. All over the world has voiced out their dissatisfaction to zionist atrocities. As a
12 wall updates, private message, poke or text, and video chat. Lastly, as a college student where we put the pros and cons of Facebook, we can conclude that, if Facebook is used in the
13    up , facebook is the link that can use to get the information fastly in one time. Everywhere , we turn today , the world is crowded with people busy looking for the jobs , or make a
14     of view about using of Facebook which is agree and disagree. Actually it is depend on how we use the facebook. If we use the facebook in a good way then it will give us more benefit .
15 In conclusion, whether we said the advantages or disadvantages of Facebook is depend on how we use the Facebook. We used in right proportions and with proper care or not. It can be most
16   nutshell, there are many pros and contras by using Facebook. It is depending on us whether we want to use it in a good way or not. Always remember to use it properly to avoid any
17 have a real meeting although it is only a meeting through internet.We also can post pictures that we want and it automatically reflects us to the others.
18 is the biggest disadvantages of Facebook. For Instance, once we stay connected to our friends, we will often log in to our Facebook account to discover what they are doing and what the
```

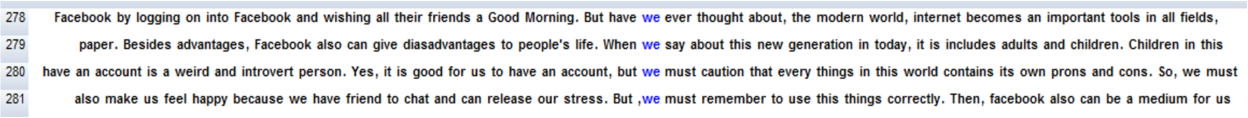
Figure 5.15: Concordance lines for *we* expressing shared neutral experiences

Hinkel (2002: pp. 53-54) argues that learners engaged in knowledge-telling deals i.e. argumentative or 'opinion' essays, focus only on two main elements, the "statement of belief and reason". It can be seen that learners express these elements by making use of shared experiences and building rapport with their audience via the various sub-types of assuming shared experiences, as described above. This, in turn, could explain the pervasive *we* in assuming shared experiences in MCSAW, as opposed to the other discourse functions. As Biber et al. (2002) state, "*we* is typical of written style, and places the focus on shared human experience or knowledge, including the speaker's" (2002: p. 96). Hence, the frequent use of modality with the personal pronoun *we* contributes to the increase of writer-visibility in Malaysian learner writing.

5.3.2.5 Emphasising or calling for the reader's attention

In MCSAW, there are only four instances (1.3%) of *we* functioning to emphasise or call for the reader's attention. According to Luzón (2009), learners may use *we* to both emphasise a claim and appeal to the reader. More specifically, use of *we* in this particular function refers to a sense of providing emphasis, in which meta-discoursal verbs comment on the text or discourse itself. However, in her study, the Spanish learners mostly used the cluster *we must say that* to imply this purpose (Luzón, 2009: p. 202). In MCSAW, this can be found in the cluster *when we say about*, which is similarly typical of spoken discourse.

The verbs used to call for the reader's attention in expert discourse tend to co-occur with adjectives such as "*important*, *interesting*, or with the modal verb *should*" (Luzón, 2009: p. 202); but instances in MCSAW were not found to be this way. In fact, *we* never co-occurs with *note*; a collocation which was also described by Luzón as characteristic of native speaker writing (e.g. *it can be noted that*). One possible explanation for this lack is the prevalence of the phrase in academic writing, which is not however common in the argumentative essay. In MCSAW, learners use *we* to emphasise a claim with verbs denoting verbal processes (*when we say*, *we must caution that*), and cognitive processes (*have we ever thought about*, *we must remember to*), as shown in Figure 5.16. In fact, the occurrences to draw the reader's attention in the learner corpus demonstrate similar expressions to the examples found in Luzón (2009), which are *we can't/mustn't forget*.



```
278   Facebook by logging on into Facebook and wishing all their friends a Good Morning. But have we ever thought about, the modern world, internet becomes an important tools in all fields,
279      paper. Besides advantages, Facebook also can give diasadvantages to people's life. When we say about this new generation in today, it is includes adults and children. Children in this
280  have an account is a weird and introvert person. Yes, it is good for us to have an account, but we must caution that every things in this world contains its own prons and cons. So, we must
281      also make us feel happy because we have friend to chat and can release our stress. But ,we must remember to use this things correctly. Then, facebook also can be a medium for us
```

Figure 5.16: Concordance lines for *we* emphasising or calling the reader's attention

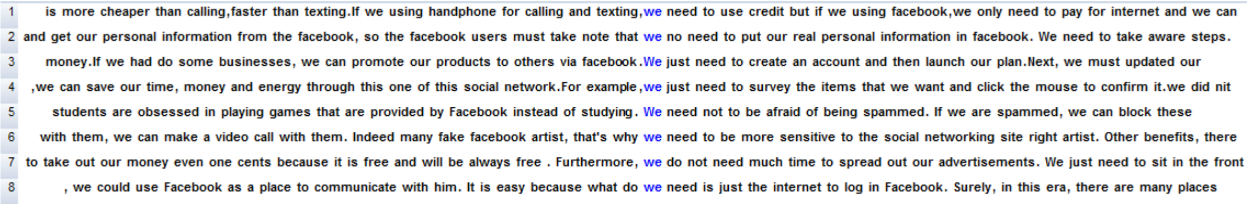
It is interesting to find that the plural pronoun *we* co-occurs with the phrasal verb *have thought* in line 278 as a question. This is described by Harwood (2005b) as formulating questions that might be posed by an imaginary readership, to enhance the interactive quality of a text. Hence, instances such as the inclusive pronoun *we* in line 278 "help to simulate reader/writer dialogism, making the reader feel involved in the argument" (ibid: p. 359). *We* also co-occurs with the verb *say* that denotes a hypothetical sense in line 279. Both examples are used to call for the reader's attention, more specifically through use of the pronoun *we* + verb of cognitive/speech; which type is not often found in the same way in native expert writing (Fawcett, 2013; Gilquin & Paquot, 2007; Luzón, 2009).

We is also found to co-occur with the modal verb *must* in phrases *we must caution that…* (line 280) and *we must remember that…* (line 281), particularly in calling the reader's attention to the following clause. In line 280, the writer warns of the 'pros and cons' of having a Facebook account, while the writer in line 281 reminds us 'to use Facebook correctly'. Furthermore, Luzón (2009) argues that occurrences of the personal plural pronoun *we* such as in the examples above do not tend to occur in formal academic writing, and hence show students' lack of awareness of the phraseology of such discourse.

5.3.2.6 Expressing opinion or volition

In identifying instances of *we* that express opinion or volition, I include *we* co-occurring with verbs of necessity (*need*), volition (*want, wish*) and deontic modality (*should, have to*) when they function to state an opinion. In MCSAW, there are 28 instances (8.9%) that were identified as denoting this function.

It is found that 8 instances of *we* co-occur with the verb of necessity *need* in expressing opinion/volition, as presented in Figure 5.17 below. There are three instances in which the negated meaning is depicted, such as *we no need to, we need not to*, and *we do not need* in lines 2, 5, and 7. In contrast to the two latter phrases, *we no need to* appears to be directly translated from the Malay equivalent, *kita tidak perlukan/memerlukan*. In fact, the two-word phrase 'no need' is often found in daily conversations, especially in colloquial Malaysian English (Manglish). Lines 3 and 4 show that *we* + *need* also co-occur with the adverb *just* in the phrase 'we just need'. *Just* is considered to be highly common in conversation, as put forth by Biber et al. (2002: p. 368): "the adverb occur[s] in conversation over 2,500 times per million words". Apart from expressing 'only' or 'no more than', Biber et al. (ibid.) state that the adverb is "also useful in focusing on the part of the clause felt to be important". Thus, for example, *just* focusses attention on the need to create an account in line 3, and the need to survey items in line 4. In the Malay language, 'we just need to' can be equally translated into *kita hanya perlukan/memerlukan*, which is suggestive of learners' L1 transfer. In line 8, assuming that the learner was aiming to use *what we need* instead of *what do we need*, it suggests a spoken-like feature of writing.

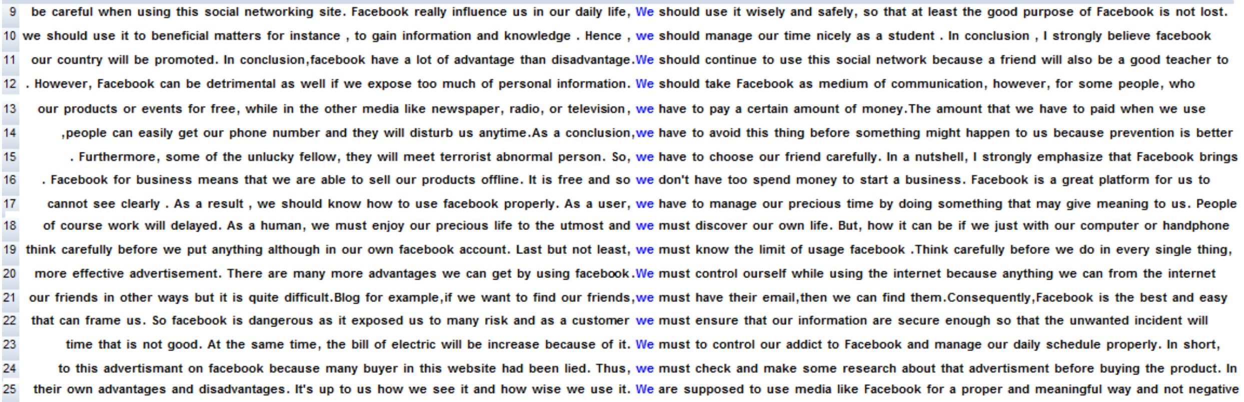


```
1    is more cheaper than calling,faster than texting.If we using handphone for calling and texting,we need to use credit but if we using facebook,we only need to pay for internet and we can
2  and get our personal information from the facebook, so the facebook users must take note that we no need to put our real personal information in facebook. We need to take aware steps.
3    money.If we had do some businesses, we can promote our products to others via facebook.We just need to create an account and then launch our plan.Next, we must updated our
4   ,we can save our time, money and energy through this one of this social network.For example,we just need to survey the items that we want and click the mouse to confirm it.we did nit
5     students are obsessed in playing games that are provided by Facebook instead of studying. We need not to be afraid of being spammed. If we are spammed, we can block these
6   with them, we can make a video call with them. Indeed many fake facebook artist, that's why we need to be more sensitive to the social networking site right artist. Other benefits, there
7  to take out our money even one cents because it is free and will be always free . Furthermore, we do not need much time to spread out our advertisements. We just need to sit in the front
8        , we could use Facebook as a place to communicate with him. It is easy because what do we need is just the internet to log in Facebook. Surely, in this era, there are many places
```

Figure 5.17: Concordance lines for *we need*

In Figure 5.18, *we* can be seen to co-occur with deontic modality *should*, *have to*, *must*, and *are supposed to* in expressing the opinion/volition meaning. As discussed in the previous chapter, modal verbs such as *must*, *have to*, *should*, *ought to*, and *need to* usually express obligation and necessity (Collins, 1991; Leech & Coates, 1980). These modals carry meanings of obligation, necessity, and requirement imposed by a source of authority. As Warner (1993) notes, the meanings of obligation and necessity that are contextually implicit in the meanings of *must*, *have to*, and *should* are determined by the speaker's assessment and decision.

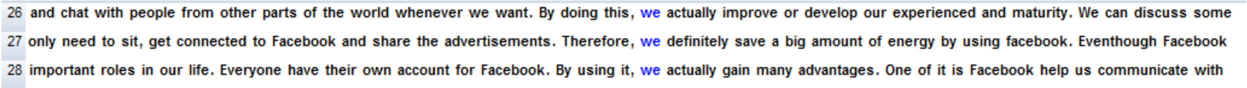
Because the pragmatic usage of modals of obligation and necessity often reflects culture-specific norms, expectations, roles, and concepts defining relationships between people and events (Sweetser, 1990), the usage of modals *must* and *have to* in these texts reflects learners' presuppositions and/or assumptions pertaining to expectations that are shared by general consensus. In other words, the examples in Figure 5.18 show the writers' opinions on what to do (e.g. *we should use it* [Facebook] *wisely/manage our time nicely*), what not to do (e.g. *we have to avoid this thing*), and what is morally or necessarily suggested of users (e.g. *we are supposed to use media like Facebook for a proper and meaningful way*).



```
9  be careful when using this social networking site. Facebook really influence us in our daily life, We should use it wisely and safely, so that at least the good purpose of Facebook is not lost.
10 we should use it to beneficial matters for instance , to gain information and knowledge . Hence , we should manage our time nicely as a student . In conclusion , I strongly believe facebook
11  our country will be promoted. In conclusion,facebook have a lot of advantage than disadvantage.We should continue to use this social network because a friend will also be a good teacher to
12 . However, Facebook can be detrimental as well if we expose too much of personal information. We should take Facebook as medium of communication, however, for some people, who
13  our products or events for free, while in the other media like newspaper, radio, or television, we have to pay a certain amount of money.The amount that we have to paid when we use
14  ,people can easily get our phone number and they will disturb us anytime.As a conclusion,we have to avoid this thing before something might happen to us because prevention is better
15  . Furthermore, some of the unlucky fellow, they will meet terrorist abnormal person. So, we have to choose our friend carefully. In a nutshell, I strongly emphasize that Facebook brings
16  . Facebook for business means that we are able to sell our products offline. It is free and so we don't have too spend money to start a business. Facebook is a great platform for us to
17  cannot see clearly . As a result , we should know how to use facebook properly. As a user, we have to manage our precious time by doing something that may give meaning to us. People
18  of course work will delayed. As a human, we must enjoy our precious life to the utmost and we must discover our own life. But, how it can be if we just with our computer or handphone
19 think carefully before we put anything although in our own facebook account. Last but not least, we must know the limit of usage facebook .Think carefully before we do in every single thing,
20  more effective advertisement. There are many more advantages we can get by using facebook.We must control ourself while using the internet because anything we can from the internet
21  our friends in other ways but it is quite difficult.Blog for example,if we want to find our friends,we must have their email,then we can find them.Consequently,Facebook is the best and easy
22  that can frame us. So facebook is dangerous as it exposed us to many risk and as a customer we must ensure that our information are secure enough so that the unwanted incident will
23  time that is not good. At the same time, the bill of electric will be increase because of it. We must to control our addict to Facebook and manage our daily schedule properly. In short,
24  to this advertismant on facebook because many buyer in this website had been lied. Thus, we must check and make some research about that advertisment before buying the product. In
25  their own advantages and disadvantages. It's up to us how we see it and how wise we use it. We are supposed to use media like Facebook for a proper and meaningful way and not negative
```

Figure 5.18: *We* co-occurring with deontic modality to express volition

Finally, it is found that 3 instances of *we* denoting opinion/volition co-occur with adverbs such as *actually* and *definitely* (shown in Figure 5.19). In lines 26 and 28, 'we *actually* + verb' signals an epistemic stance, where the examples comment on the reality or actuality of the proposition: the improvement of experience (line 26), and gaining many advantages (line 28). The Malay equivalent for *actually*, which is *sebenarnya*, also expresses the same epistemic stance. Contrary to the common use of *sebenarnya* following the personal pronoun in the Malay language (*kita sebenarnya*…), in English this is deemed unusual (Gablasova & Brezina, 2015). In fact, this syntactical pattern is not found in the reference language variety, and thus may suggest learners' direct translation of the Malay language. The adverb *definitely*, on the other hand, is seen to be used in line 27 to express the speaker's or writer's emphasis towards the proposition: the prospect of saving energy via Facebook. Similar to the use of *actually* following immediately after the pronoun *we*, 'we definitely' can also be found in a Malay equivalent, as *kita sememangnya*…. In sum, the distinctive use of 'we actually/definitely' as described above can be argued as being the result of Malaysian learners' influence from the first language, Malay, in their writing.



```
26 and chat with people from other parts of the world whenever we want. By doing this, we actually improve or develop our experienced and maturity. We can discuss some
27 only need to sit, get connected to Facebook and share the advertisements. Therefore, we definitely save a big amount of energy by using facebook. Eventhough Facebook
28 important roles in our life. Everyone have their own account for Facebook. By using it, we actually gain many advantages. One of it is Facebook help us communicate with
```

Figure 5.19: Concordance lines for remaining instances of *we* denoting volition

Meanwhile, the use of *we* in expressing opinion or volition in learner writing may be due to the nature of the argumentative essay: writers are encouraged to demonstrate the relevance of the claim to their proposition (Hyland, 1990). This, as Hyland remarks, is a feature of persuasion in the argument stage, "declaring opinion, which aims for maximum effect with minimum regard for opposing views" (1990: p. 73), but is often at the expense of adherence to the academic writing genre.

5.4 Summary

This chapter analysed Malaysian learners' use of the personal pronoun *we*, found to be prevalent in MCSAW, given results of the keywords analysis in Chapter 4. Similar to the analysis of *can*, *we* was investigated in terms of range, dispersion and collocation by using WordSmith Tools and the reference corpus, LOCNESS.

It was found that the personal pronoun *we* is proportionately distributed across the three major L1 groups in MCSAW, with only 4% difference in the Malay texts. Furthermore, *we* is mostly dispersed in the middle and end part of texts in MCSAW, suggesting that it is used with particular discourse functions. These functions were examined in Section 5.3.2 of this chapter, and are summarised below.

In contrast to the salient use of *can*, *we* shared eleven similar collocates in both MCSAW and LOCNESS. This suggests that *we* is used in similar ways by Malaysian learners and in the reference language variety. Both groups of novice writers produced expressions of modality with the use of *we must* and *we need* constructions, and employed some form of 'solidarity strategy' through use of inclusive *we*, particularly with cognitive and speech verbs such as *we know*, *we can see*, *we can ask*, and *we say*. Use of reflexive pronoun *ourselves* and adverbs *today* and *already* can also be seen in both corpora, suggesting that the use of inclusive *we* contributes to a high degree of writer/reader visibility, as mentioned by previous research (Cobb, 2003; McCrostie, 2008; Petch-Tyson, 1998). Several explanations for these features in learner writing include the influence of spoken language, L1 transfer, aspects of teaching, and cultural factors (Petch-Tyson, 1998; Gilquin & Paquot, 2008; Paquot, 2010). However, L1 transfer would not apply to the native speakers of LOCNESS who also used these features.

Further qualitative analysis was carried out by examining concordance lines for the personal pronoun *we* in MCSAW. More specifically, this thesis attempted to determine how Malaysian learners use the personal pronoun *we*, by adapting Luzón's (2009) classification of discourse functions. The findings revealed that *we* in MCSAW is mostly used in assuming shared experiences/knowledge, goals, or beliefs (86.3%), followed by expressing opinion or volition (8.9%). Learners used *we* less in contexts that involve stating conclusions (1.6%), emphasising or calling for the readers' attention (1.3%), stating the thesis (1.0%), and guiding the reader through the text (1.0%). This indicates that many Malaysian learners used *we* as an inclusive first-person pronoun that refers to a generic substitute representing a large group of people, particularly to assume shared knowledge (e.g. as *we know*, *as we all know*) and to express volition (e.g. *we should*, *we have to*). This may explain the dispersion of *we* in middle/end parts of argumentative essays. Learners under-used *we* in guiding the reader through the text and calling for reader's attention, possibly due to limited knowledge of these specific discourse functions in writing. Where attempts were made to produce these functions, learners' writing style appeared as more conversational.

The findings also showed several instances of redundancy or tautology in the use of *we*, in which most *we* instances could be replaced with the passive voice or inanimate subjects. Repeated lines that were found in more than one essay suggest plagiarism or the possible influence of prompts that were used in the classroom. It was also found that *we* co-occurs with high-frequency common verbs such as *get*, *have* and *know*, contributing to learners' writing being more spoken-like. Learners were also found to demonstrate direct translation of their first language, Malay, in many *we* occurrences, namely: *we no need to* ('kita tidak perlukan/memerlukan'); *we just need to* ('kita hanya perlukan/memerlukan'); *we actually* ('kita sebenarnya'); and *we definitely* ('kita sememangnya').

In other words, non-native writers in MCSAW are highly influenced by acquiring chunks of language in everyday conversation, which in turn affects their language use in writing. More specifically, the highly salient combination *we* + *can* was found most prevalent in expressing shared experiences, which is commonly found in spoken discourse. In addition to the analysis of *can* in Chapter 4, use of *we can* in the present chapter revealed the incorporation of inclusive *we* with the dynamic modality sense of *can*, particularly in expressing assumed knowledge related to the essay topics. This confirms Hinkel's (2009: p. 680) claim that essay topics potentially affect learners' use of modal verbs. He implies that one of the pitfalls of broad-based topics is that "reliance on one's own experience and knowledge in lieu of factual or demonstrable evidence can lead to greater cultural boundedness and personalization of writing" (ibid: p. 681). Hence, in many cases of the phrase in MCSAW, it can be argued that this particular style in Malaysian learner writing must have contributed greatly to the description of higher visibility of the author and the higher use of spoken language features, i.e. interpersonal writing that is not favoured in many forms of academic writing, including the argumentative essay.

One possible conclusion for the overuse of *we* is that Malaysian learners share the same problem with many ESL and EFL learners: their writing is highly interpersonal, usually aiming to show solidarity and to engage with readers of the text (McCrostie, 2008). When learners write in their argumentative essays, they were found to be mostly influenced by or sub-consciously applying their conversational speech into writing. Nevertheless, native-speaking writers in the reference language variety were also found to use substantial amounts of the inclusive pronoun *we* in their writing. Some similarities were found, despite differences in the frequency of use. To summarise, although the personal pronoun *we* can be used to perform rhetorical functions in writing, many learners find it difficult, as combining both formal and informal features is more challenging than following a strictly formal style (Chang & Swales, 1999; McCrostie, 2008). One particular difference is that learners primarily draw on shared experience in the use of the pronoun *we*, and where attempt is made to perform rhetorical strategies with *we* as indicating authorial stance, instances showed idiosyncratic usage.

Chapter 6: Key Lexical Bundles

6.1 Introduction

In the previous two chapters, analyses based on individual keywords were considered for the discovery of the salient use of modal verb *can* and personal plural pronoun *we* in Malaysian learner writing. Another way to investigate differences is by analysing the most frequently recurring sequences of words, i.e. lexical bundles.[54] As mentioned in Chapter 2, lexical bundles have been investigated in numerous learner corpus studies, to analyse lexico-grammatical features of language for a variety of communicative types and purposes (Biber, 2006; Cortes, 2004; De Cock, 1998; Hyland, 2008a, 2008b). It has also been mentioned that the ability to understand and use lexical bundles appropriately are key to native-like fluency (Simpson, 2004: p. 37). According to Wray (1999: p. 225), the absence of lexical bundles in learners' discourse may result in unidiomatic-sounding writing style. As outlined in Chapter 3, the present chapter presents results for lexical bundle analyses of the Malaysian corpus (MCSAW) against its reference language variety, LOCNESS. Key lexical bundles are firstly investigated in terms of their frequencies as well as distribution in both corpora. This is discussed in two sections: key 4-word bundles (Section 6.2), and key 3-word bundles (Section 6.4). The larger sequences are discussed before the smaller sequences to avoid repetition, since 3-word bundles are often integrated in the larger 4-word bundles (Cortes, 2004: p. 401). Following this, key lexical bundles are further categorised and analysed according to their functions, including qualitative analysis of the most recurrent bundles by examination of concordance lines.

6.2 Key 4-word lexical bundles

As discussed in Chapter 3, lexical bundles are analysed using WordSmith Tools, which compare 3-word and 4-word bundles in the Malaysian learner corpus and reference corpus. The first analysis examines key 4-word lexical bundles in MCSAW, limited to those that occur with a minimum raw frequency of one in LOCNESS to allow comparison of usage. Consideration of bundle lengths was decided following arguments made by Biber et al. (2002)

[54] It is worth restating that different terms such as formulaic sequences, multi-word units, clusters, and n-grams are largely equivalent; and this chapter will refer to them as lexical bundles for ease of comprehension.

and Cortes (2004: p. 401), namely that 4-word lexical bundles are more frequent than 5-word bundles, and thus "present a wider variety of structures and functions to analyse". Longer recurrent sequences, i.e. 5-word and 6-word bundles can be found, but they are much less common (Biber et al., 2002). However, 3-word lexical bundles are also investigated to discern whether they encompass longer, i.e. 4-word lexical bundles as claimed in Cortes (2004). More specifically, initial description of key bundles is based on 3-word and 4-word sequences, but only *frequent* 3-word and 4-word sequences are considered in the more detailed (qualitative) analyses.

Table 6.1 presents, in descending order of relative frequency in LOCNESS, a list of 4-word bundles, which are statistically more significant in MCSAW, compared against LOCNESS: that is, they are 'key' bundles. Similar to the keyword analyses in Chapter 4 and Chapter 5, an analysis of key bundles is useful as a starting point in investigating learners' use of lexical bundles contrasted with a reference language variety, following the corpus-driven approach. As can be seen, 26 key lexical bundles[55] in MCSAW have at least one occurrence in LOCNESS. For example, the highest-ranked bundle in the list is *is one of the*, which occurs 130 times in MCSAW and 30 times in LOCNESS, while the least-ranked bundle, *us the opportunity to*, appears 82 times in the learner corpus and only once in the reference language variety. According to Cortes (2004: p. 401), "many four-word bundles hold three-word bundles in their structures (as in *as a result of*, which contains *as a result*)". Given this, findings from Table 6.1 will be compared with shared 3-word lexical bundles (see Section 6.4) in order to confirm this.

[55] 'Shared lexical bundles' will be used as a short-hand from here on, to refer to those lexical bundles that are 'key' in MCSAW *and* occur at least once in the reference corpus. Unshared lexical bundles are not investigated in this chapter, because the focus is on comparing the use of bundles in both MCSAW and the reference language variety. This will include qualitative analysis of lexical bundles as used in both corpora.

Table 6.1: Shared 4-word lexical bundles that occur at least once in LOCNESS

N	Key word	Freq.	%	Texts	RC. Freq.	Keyness
1	is one of the	130	0.07	104	30	127.28
2	all over the world	57	0.03	46	11	61.27
3	the best way to	41	0.02	35	9	41.26
4	is the best way	31	0.02	31	4	39.30
5	one of the biggest	26	0.01	26	3	34.20
6	of the most popular	26	0.01	26	3	34.20
7	anywhere in the world	63	0.03	61	3	101.16
8	there are many advantages	36	0.02	32	2	56.35
9	with the help of	43	0.02	43	2	69.29
10	people around the world	123	0.06	107	2	221.01
11	this is the best	22	0.01	22	1	35.57
12	of advantages and disadvantages	23	0.01	21	1	37.44
13	most of the people	25	0.01	24	1	41.17
14	a lot of time	26	0.01	24	1	43.04
15	there are a lot	28	0.01	24	1	46.79
16	are a lot of	29	0.01	25	1	48.67
17	can help us to	30	0.02	23	1	50.55
18	it is the best	31	0.02	31	1	52.43
19	people in the world	35	0.02	31	1	59.99
20	have a lot of	37	0.02	30	1	63.77
21	all around the world	49	0.02	37	1	86.59
22	in many ways and	50	0.03	50	1	88.50
23	important part of our	51	0.03	51	1	90.41
24	to know more about	58	0.03	57	1	103.79
25	one of the best	74	0.04	69	1	134.48
26	us the opportunity to	82	0.04	63	1	149.86

Out of the 26 shared lexical bundles, 7 are highly distributed across more than 50 texts in the Malaysian corpus. *People around the world* occurs in a majority of texts (107), followed by *is one of the* (104 texts). Other bundles that are approximately evenly distributed in the corpus include *one of the best* (69), *us the opportunity to* (63), *anywhere in the world* (61), *to know more about* (57), and *important part of our* (51). These recurrent bundles also show high keyness values, which means their difference of occurrence between the two corpora is statistically more significant. More importantly, as argued by Hyland (2012), their recurrence in multiple texts suggests at least some perceptual salience among users, and thus a particular writing style. Bundles that show a huge difference in terms of relative frequencies and occur more widespread in MCSAW, consequently, may indicate Malaysian learners' overuse or idiosyncratic uses of these bundles. However, qualitative analysis will be undertaken in Section 6.5 to determine if such bundles can indeed be interpreted as characteristic of Malaysian learner writing style or if other explanations are more suitable.

Another observation shows that, although lexical bundles are usually not complete grammatical units, they tend to have particular grammatical characteristics (Biber et al., 2002: p. 445). For example, shared 4-word bundles can be grouped into 10 bundles incorporating the verb phrase[56] (VP) (*is one of the*, *is the best way*, *there are many advantages*, *this is the best*, *are a lot of, have a lot of*, *there are a lot*, *can help us to*, *it is the best*, *to know more about*), 9 bundles incorporating the noun phrase (NP) (*the best way to*, *one of the biggest*, *a lot of time*, *important part of our*, *one of the best*, *people around the world*, *most of the people*, *people in the world*, *us the opportunity to*), and 7 bundles incorporating the prepositional phrase (PP) (*all over the world*, *of the most popular*, *anywhere in the world*, *with the help of*, *of advantages and disadvantages*, *all around the world*, *in many ways and*). This shows that learners make use of more 4-word VPs, followed by 4-word NPs and 4-word PPs. In his investigation on bundles in speech and writing, Biber (2009: p. 300) found that these tendencies are more characteristic of speech. Academic prose however, contains more NP-/PP-based bundles.

It is also found that most of the 4-word bundles consist of words that evaluate or qualify a specific entity (e.g. *best*, *biggest*, *most*, *advantages*); and in turn, indicate bundles that are related to the genre and type of argumentative essay in MCSAW, where the prompt asks learners to discuss and evaluate the advantages/disadvantages of Facebook and living in a hostel. Other bundles show the use of closed-class/function words (e.g. *can, the*, *most*) and high-frequency common words (e.g. *way*, *help*, *have*), which have been expressed by Lee and Chen (2009: p. 286) as being "not specific to academia, but of more general currency". The highly frequent use of function and simple words possibly suggests that they may be constructed in recurrent patterns that are idiosyncratic rather than being randomly used. The less-varied bundles also indicate learners' limited vocabulary. Essentially, the 4-word lexical bundles presented in Table 6.1 show that they are not frequently shared in LOCNESS, but are over-used in MCSAW. This, then, points to the interest that lies in exploring types of lexical bundles in order to study differences between functional uses in learner writing compared to the reference language variety, conducted in Section 6.5. Before doing so, the next section will attempt to categorise these bundles functionally, based on previous research into types of bundles in academic discourse.

[56] Following Chen and Baker (2010: p. 34), NP- based bundles include any noun phrases with post-modifier fragments, such as *the role of the* or *the way in which*. PP-based bundles refer to those starting with a preposition plus a noun-phrase fragment, such as *at the end of* or *in relation to the*; and with regard to VP-based bundles, any word combinations with a verb component, such as *in order to make* or *was one of the*.

6.3 Functional categorisation of lexical bundles

Researchers who work on academic discourse have proposed different ways of categorising bundles. For example, Hyland (2008a, 2008b) has categorised bundles into three main groups, namely, Research-oriented bundles (description of research experiences), Text-oriented bundles (organisation of the text/argument), and Participant-oriented bundles (writer/reader-focused features of the discourse). For the most part, Hyland identified research-oriented bundles that consist of bundles indicating time/place (*at the beginning of*, *in the present study*), process demonstration (e.g. *the purpose of the, the operation of the*), quantification (e.g. *the magnitude of the*, *a wide range of*), description (*the structure of the*, *the size of the*), and topic related to the field of research (*in the Hong Kong, the currency board system*). In his works (Hyland 2008a, 2008b), lexical bundles were investigated in terms of their frequencies and uses in research articles, PhD dissertations and MA/MSc theses, from four disciplines. Unlike academic essays, that are research-oriented and focus on process and results of research such as described in Hyland's studies, argumentative writing is often more opinionated and topic related (Johns, 1997). More specifically, argumentative essays would not contain the elements of research that are depicted in the first category mentioned by Hyland.

Similarly, Biber et al. (2004) classified common lexical bundles into three types: Stance bundles, Discourse Organising bundles, and Referential bundles. Stance bundles are identified as bundles "express[ing] attitudes or assessments of certainty that frame some other proposition" (Biber et al., 2004: p. 384); Discourse organisers are bundles that "reflect relationships between prior and coming discourse" (Biber et al., 2004: p. 384); while Referential bundles indicate some form of "direct reference to physical or abstract entities, or to the textual context itself, either to identify the entity or to single out some particular attribute of the entity as especially important" (Biber et al., 2004: p. 384). According to their taxonomy, Stance bundles are further comprised of two types: 'Epistemic stance' (e.g. *I don't know what, the fact that the*) and 'Attitudinal/modality stance' (e.g. *I don't want to*, *it is important to*). Discourse organisers are also sub-divided into two types, namely 'Topic introduction/focus' (e.g. *if we look at, I would like to*) and 'Topic elaboration/clarification' (e.g. *I mean you know, on the other hand*); while Referential expressions comprise 'Identification/focus' (e.g. *is one of the, one of the most*), 'Imprecision' (e.g. *or something like that, and things like that*), 'Specification of attributes' (e.g. *there's a lot of*, *as a result of*), and 'Time/place/text reference' (e.g. *in the United States, at the end of*). Contrary to Hyland's classification (2008a, 2008b), which focusses more on research-based genres and sub-categories that “specifically reflect the concerns of research writing” (Hyland, 2008b: p. 13), Biber et al.’s (2004: p. 383) taxonomy extends from their previous research on lexical bundles developed for conversation and academic prose (Biber et al., 2003), and thus is deemed to be more generic.

More precisely, Biber et al.’s (2004) study explored academic language used in university classroom teaching and textbooks represented in the U.S., and “outlined a taxonomy of the major discourse functions served by lexical bundles […] that can potentially be realised in any register” (ibid: p. 396). Their findings revealed that all three types of bundles are commonly found in classroom teaching, with the preference for referential bundles being extremely common in classroom teaching and less common in textbooks and academic prose. This is, therefore, relevant in examining types of lexical bundles found in learner argumentative writing such as in MCSAW, in which texts in the corpus are written essays produced in a classroom setting at a given time.

Chen and Baker (2014, 2010) build on Biber et al.’s (2004) functional categorisation, but modify this categorisation scheme slightly. They assert that not only do discourse organisers introduce or elaborate texts but they also include bundles that show/make inferences (e.g. *in the sense that, as a result of*). Their study looked at lexical bundles found in native (expert and peer) and learner academic writing, and concluded that published academic writing was found to exhibit the widest range of lexical bundles, whereas L2 student writing showed the smallest range. Of these three categorisation schemes, Biber et al.’s (2004) taxonomy appears to be the most relevant for the present analysis of lexical bundles in MCSAW, and therefore will be used as a starting point for the new categorisation scheme adopted and adapted in this chapter. Furthermore, it is also important to incorporate some of Chen and Baker’s (2014, 2010) suggestions, as well as developing new sub-categories, which arose from the need to classify the identified bundles more precisely.

Table 6.2 presents the categorisation scheme used for analysing bundles found in the Malaysian learner corpus, with some examples (both 3- and 4-word bundles) for each category provided from MCSAW. Categories and sub-categories that are new are symbolised with an *, while the symbol † identifies categories that are taken from Chen and Baker (2014, 2010). Categories without a symbol were adopted from Biber et al. (2004), as explained above.

Table 6.2: Functional categorisation of lexical bundles

Categories	Definition	Examples
Referential bundles	These refer to physical, abstract or contextual aspects, including those that focus on a particular feature of an entity.	**6.1.1 Identification/focus bundles** – to identify or focus on the phrase following the bundle, including existential *there* constructions† (e.g. *is one of the*, *there are some*[57]) **6.1.2 Bundles specifying attributes of following nouns/entities** – to identify specific attributes/qualities (including quantities) of the following head noun/entity 6.1.2.1 Not incorporating the specified entity* – *have a lot of, the popularity of* 6.1.2.2 Incorporating the specified entity* – *a lot of time* **6.1.3 Bundles specifying attributes of preceding nouns/entities*** **–** to identify specific attributes/qualities of the preceding head noun/entity 6.1.3.1 Not incorporating the specified entity* – *of the most popular*, *of our life* 6.1.3.2 Incorporating the specified entity* – *people around the world*[58] **6.1.4 Time/place/text-deixis bundles**[59] – referring to particular places, times, or locations (e.g. *all over the world, in our life*) **6.1.5 Imprecision bundles** – to indicate that a specified reference is not necessarily exact, or to indicate that there are additional references of the same type that could be provided (e.g. *in many ways and, and so on*) **6.1.6 Other referential bundles*** – bundles that make reference to physical or abstract entities or processes and are often topic-related; several are negated; includes adverbials (e.g. *low income families*, *face to face*, *students do not, we do not; as a student*)

[57] Bundles with *there* constructions could also have been included as 'discourse organisers', but in order to avoid double-classification, I follow Chen and Baker (2014: pp. 19-23), in which *there are some* "qualify the proposition expressions related to something potentially gaugeable in terms of size, amount, extent, and so on"; often appearing in existential *there* constructions.

[58] Since these bundles contain a place reference, they could also be identified as place deixis, but occur here with a preceding noun.

[59] Some of these could also be identified as bundles specifying attributes of preceding entities (e.g. *in our life, of all time*), but were classified as time/place/text-deixis here because they contain some reference to time and place.

Discourse organising bundles	These bundles are concerned with the topic introduction/focus, elaboration/clarification and inference.	**6.2.1 Topic introduction/focus bundles** – expressions of beginning a topic (e.g. *first of all*) **6.2.2 Topic elaboration/clarification bundles** – relates to additional explanation or clarification, usually of the subject (e.g. *is free and*, *is the most*) **6.2.3 Inferential bundles†** – making inferences (e.g. *as a conclusion*, *this is because*)
Stance bundles	These provide a frame for the interpretation of the following proposition, conveying two major kinds of meaning: epistemic and attitude/modality.	**6.3.1 Epistemic stance bundles** comment on the knowledge status of the information in the following proposition: certain, uncertain, or possible; and make reference to status of information, e.g. as opinion, knowledge* (e.g. *in my opinion*, *as we know*) **6.3.2 Attitudinal/modality stance bundles** express attitudes (self or other) towards the actions or events, usually in distinguishing desire, directives, intention/prediction, ability, importance and emotivity 6.3.2.1 Desire – bundles with *want, like, decide* (e.g. *we want to*) 6.3.2.2 Obligation/directive – bundles with *have to, should, need to* (e.g. *we have to*) 6.3.2.3 Ability – all bundles incorporating *can* and words referring to opportunity, chance, help (e.g. *with the help of*, *us the opportunity to*, *you can use*) 6.3.2.4 Importance* – bundles showing significance (e.g. *important part of our*) 6.3.2.5 Emotivity* – bundles showing an assessment of an entity or proposition as 'good/bad', including bundles incorporating the word forms *best, advantage, pros and cons*, etc. (e.g. *is the best way, the advantages of*)

The first category (**Referential bundles**) classifies lexical bundles that refer to physical, abstract or contextual aspects, including those that focus on a particular feature of an entity, and are further sub-categorised into six types. Identification/focus bundles, which focus on the noun phrase following the bundle (e.g. *is one of the + noun/NP*), is a category adopted from Biber et al. (2004), but includes existential 'there' constructions, following Chen and Baker (2014, 2010) such as the bundle *there are some*. These bundles usually identify an entity that follows the bundle, or pinpoint this entity as especially important. Bundles specifying attributes identify specific qualities of the following nouns/entities (Biber et al. 2004), but additionally of the preceding nouns/entities (including quantities). They are further classified into two new sub-categories: those that incorporate the specified entities (e.g. *a lot of* <u>*time*</u>, <u>*people*</u> *around the world*), and those that do not incorporate the specified entities, such as *have a lot of,* and *of the most popular*. Also adopting from Biber et al. (2004), time/place-text-deixis mainly show or make reference to particular time, place, or locations (e.g. *all over the world*, *in our life*); while imprecision bundles refer to bundles that are vague or that indicate imprecise reference (e.g. *in many ways and*, *and so on*). Finally, I have added a new 'catch-all' category called 'other referential bundles', which includes all other bundles that make reference to physical or abstract entities or processes (including adverbials), but are not instances of the above-mentioned types. Mostly, these bundles are topic related (e.g. *low income families*, *face to face*), and several are negated (e.g. *students do not*, *we do not*).

As suggested by Biber et al. (2004) **Discourse organising bundles** or discourse organisers function to signal readers of the writer's intention – either to introduce or elaborate – on a subject matter. Such examples include topic introduction bundle *first of all*, and topic elaboration bundle *is free and*. These two categories are taken from Biber et al. (2004), while the present categorisation scheme additionally includes Chen and Baker's (2014; 2010) inferential bundles, such as *as a conclusion*, and *this is because*.

The final category: **Stance bundles**, presents lexical bundles that demonstrate the writer's comment on the knowledge status of a proposition as being certain, uncertain or probable/possible (i.e. epistemic stance bundles), as well as expressing the writer's attitudes towards actions or events described in a proposition (i.e. attitudinal/modality stance bundles). Stance bundles can be personal or impersonal, as stated by Biber et al. (2004: p. 389). Personal stance bundles include those that are attributed to the speaker/writer (e.g. *and I think that*), whereas impersonal stance bundles express similar meaning without being attributed directly to the speaker/writer (e.g. *are more likely to*). Contrary to Biber et al. (2004), who limit stance bundles to those bundles expressing speaker/writer stance, these bundles may make reference to the speaker's own stance (e.g. ***we*** *want to*) as well as to the stance of addressees or third parties (e.g. ***you*** *want to,* ***they*** *don't like*) (Bednarek, 2008a: p. 21). Furthermore, in my categorisation scheme, epistemic stance bundles also comprise reference to status of information, for instance as knowledge or opinion (e.g. *in my opinion*), including bundles incorporating the verb *know* itself (e.g. *to know more about*, and *as we know*).

Attitudinal/modality stance bundles, on the other hand, are divided into six sub-categories. Desire bundles, which express wishes/desires/wants, include bundles with *want*, *like* and *decide* (e.g. *we want to*, *they don't like*, *have decided to*); while obligation/directive bundles, which direct the listener/reader to do something, include bundles with *have to*, *should* and *need to* (*we have to*, *should be given*, *just need to*). Ability bundles include all bundles incorporating *can* and words referring to opportunity, chance, help etc. (e.g. *can help us to*, *you can use*); while importance bundles comprise those that "evaluate the world (and discourse about it) according to the speaker's subjective evaluation of its status in terms of importance, relevance and significance" (Bednarek, 2008a: p. 16) (e.g. *important part of our*). Finally, emotivity bundles are bundles that express "the writer's evaluation of aspects of events as good or bad, i.e. with the expression of writer approval or disapproval" (Bednarek, 2008a: p. 15), including bundles with the word forms *best*, *advantage*, *pros and cons*, etc. (e.g. *is the best way*, *the advantages of*).

Table 6.3 presents the 26 shared 4-word bundles according to the broad classification (Referential, Discourse Organising, or Stance), as described in Table 6.2. There are 14 bundles which are identified as Referential, 12 as Stance bundles,[60] while none are characteristic of Discourse Organising bundles. Contrary to Chen and Baker's (2010: p. 39) findings that student texts contained far more discourse organisers than the reference language variety, this first description of types of bundles in MCSAW appears to suggest otherwise. One possible explanation would be in the different ways bundles are categorised: discourse organisers in Chen and Baker's (2014, 2010) studies include bundles with lexis that denote sense of importance (e.g. *is more important than*, *a very important role*), and emotivity (e.g. *it is a good*, *the best way to*), whereas in the present study, such bundles are identified as Stance (as mentioned in Table 6.2).[61] It could be argued that bundles containing adjectives such as

[60] Certain of these bundles could also be classified differently, for instance as identification/focus bundles (e.g. *one of the best*), as bundles specifying attributes of preceding entities (e.g. *of advantages and disadvantages*), as topic introduction (e.g. *there are many advantages*), or as topic elaboration/clarification (e.g. *is the best way*); but were classified as Emotivity because they are ultimately seen as expressing opinion through use of lexis such as *best*, *advantages*, etc. The same applies to the classification of relevant 3-word bundles such as *the advantages of*, which are classified as stance but also specify the attributes of an ensuing entity.

[61] It is important to highlight that the types of sequences included in Chen and Baker (2014, 2010) are also different because they investigated and compared bundles in learner writing to expert writing (published academic writing), which does not make the results directly comparable. In addition, they use a different frequency threshold than that used in this thesis.

important, *different*, *best*, *good*, and adverbs such as *more*, *very*, *most*, are employed mostly as overt expressions of personal attitudes or feelings towards the content of a clause.

Table 6.3: Classification of shared 4-word lexical bundles

N	Key word	Freq.	Category	Texts	RC. Freq.	Keyness
1	is one of the	130	Referential	104	30	127.28
2	all over the world	57	Referential	46	11	61.27
3	the best way to	41	Stance	35	9	41.26
4	is the best way	31	Stance	31	4	39.30
5	one of the biggest	26	Referential	26	3	34.20
6	of the most popular	26	Referential	26	3	34.20
7	anywhere in the world	63	Referential	61	3	101.16
8	there are many advantages	36	Stance	32	2	56.35
9	with the help of	43	Stance	43	2	69.29
10	people around the world	123	Referential	107	2	221.01
11	this is the best	22	Stance	22	1	35.57
12	of advantages and disadvantages	23	Stance	21	1	37.44
13	most of the people	25	Referential	24	1	41.17
14	a lot of time	26	Referential	24	1	43.04
15	there are a lot	28	Referential	24	1	46.79
16	are a lot of	29	Referential	25	1	48.67
17	can help us to	30	Stance	23	1	50.55
18	it is the best	31	Stance	31	1	52.43
19	people in the world	35	Referential	31	1	59.99
20	have a lot of	37	Referential	30	1	63.77
21	all around the world	49	Referential	37	1	86.59
22	in many ways and	50	Referential	50	1	88.50
23	important part of our	51	Stance	51	1	90.41
24	to know more about	58	Stance	57	1	103.79
25	one of the best	74	Stance	69	1	134.48
26	us the opportunity to	82	Stance	63	1	149.86

Most shared 4-word bundles in Table 6.3 are identified as referential, including the first-ranked (i.e. most frequent in terms of relative frequency) bundle *is one of the*, and the most statistically significant bundle *people around the world*. Such bundles, as previously mentioned, introduce a particular subject as particularly important, such as in *Facebook* *is one of the*, or identify entities with specific attributes, as in *most of the people* *who*. Bundles that specify attributes to these entities include quantifying of the indicated/non-indicated entities (e.g. *a lot of time*, *are a lot of*), and framing[62] (i.e. used to specify a given attribute or condition) of the indicated/non-

[62] This is following Chen and Baker (2014, 2010) on the sub-category of Referential bundles, which is used to specify a particular attribute of an entity or condition (e.g., *in terms of the*, *in the context of*, *the nature of the*,

indicated entities (e.g. *people around the world*, *of the most popular*).[63] Furthermore, these referential-type bundles are found to be mostly topic-oriented. For example, in the three instances below, bundles that are underlined all make reference to Facebook:

Most of the people have more than one Facebook account nowadays (185txt_MCSAW)

We will spend *a lot of time* in front of the computer for Facebook without us realize it (258txt_MCSAW)

It is clear that, using Facebook *have a lot of* advantages and disadvantages (310txt_MCSAW)

This also appears to be the case in LOCNESS. The referential bundle *is one of the* is found to be used in relation to the topics in LOCNESS (e.g. *Boxing is one of the most popular sports of this era*; *Genetics is one of the fastest growing fields of science in the world today*). Regardless, the bundle is over-used in MCSAW because it occurs over four times more frequently than in LOCNESS. Previous studies on lexical bundles argue that *is one of the* is among the most frequent shared bundle in novice academic writing (Ädel & Erman, 2012; Chen & Baker, 2014), and Biber et al. (2004) describe it as typical in conversation. Similarly, the existential *there* construction bundle *there are a lot* is over-used, and appears to be topic-related since there are many uses of this bundle as regards the advantage/disadvantages of using Facebook. Chen and Baker (2014: pp. 17-18) argue that “the prevalence of copula *be* constructions [as in the bundles *is one of the, there are a lot*,] in learner writing […] conforms to the norm of conversation rather than that of the written register”.

Other Referential bundles include those indicating vagueness, i.e. imprecision bundles – *in many ways and* – and those that refer to particular places or time (*all over the world/anywhere in the world/all around the world*). Arguably, these bundles are also found to contain reference to the topic of essays, such as in *It [Facebook] is helping us in many ways and also harming us in other ways* (MCSAW_26.txt) and *This give benefit to us as we can know more about*

the existence of a), and is characteristic of academic writing. In other words, bundles that specify attributes to the entity are also considered framing bundles.

[63] Given the adjective *popular,* this bundle could alternatively be classified as referring to the attitudinal stance of third parties (something that is popular would also be liked by many).

people around the world with Facebook (MCSAW_16.txt). Topic dependency will be explored further in the qualitative analysis in Section 6.5.

Moving on to stance bundles, it is found that there is only one epistemic 4-word stance bundle in Table 6.3, which is *to know more about*, while the remaining 4-word stance bundles are recognised as attitudinal. It must be noted here that the bundle is only 'epistemic' (or 'evidential') in as far as it clearly implies that there is not enough knowledge about X. As noted above, I have taken a broad approach to classifying a bundle as 'epistemic stance' in this sense. In other words, no 4-word stance bundles express the writer's comment on the knowledge status of a proposition as being certain, uncertain or probable/possible.

Instead, attitudinal stance bundles were found to be the most frequent type of stance bundles, perhaps unsurprisingly given the argumentative nature of this genre (11 bundles: *the best way to*, *is the best way*, *there are many advantages*, *with the help of*, *this is the best*, *of advantages and disadvantages*, *can help us to*, *it is the best*, *important part of our*, *one of the best*, and *us the opportunity to*). One observation can be made as regards the frequent use of superlatives, which shows writers' opinions towards a particular proposition such as bundles with the word *best*, which express emotivity (*the best way to, is the best way*, *this is the best, it is the best, one of the best*). For example, in *This is the best way to find friends in school,* and *Facebook is the best way to communicate,* the writers demonstrate a direct and strong assertion of his/her position on what they believe to be 'the best' way of finding friends in school and source of communication.

In addition, attitudinal stance bundles can be seen to incorporate essay prompts *advantages* and *disadvantages* (*there are many advantages*, *of advantages and disadvantages*), and therefore can also be argued to be topic-related: for example, *I strongly emphasize that Facebook brings lots of advantages and disadvantages*/ *There are many advantages of using Facebook.*

The bundle *important part of our,* which is sub-categorised as bundles expressing importance, is also seen to relate to the topic, such as in *Facebook has become a very important part of our lives*. This shows learners' attitude or feelings about the entity Facebook as being important, significant or necessary in 'our' lives, and that the use of the inclusive pronoun *our* appeals to the strategy of shared experiences with readers. The remaining bundles are suggestive of expressing 'Ability', given the incorporation of the modal verb *can* (*can help us* *to*) and words referring to opportunity, chance, and/or help (*with the help of*, *us the opportunity to*). Examples include *Facebook can help us to connect to different people from anywhere in the world*, and *With the help of Facebook we can connect to different people from anywhere in the world*. A reason for the highly frequent use of bundles that express ability could be the learners' overuse of the modal verb *can*, which has already been discussed in Chapter 4.

Interestingly, as already mentioned above, there are no examples of Discourse Organising bundles found in the list. While such bundles may occur in MCSAW, they are not over-used when compared to the reference learner variety. The examination of shared 3-word bundles will identify whether Discourse Organising bundles are found to be more prevalent in shorter strings of words.

6.4 Key 3-word lexical bundles

Similar to the previous analysis of 4-word bundles, 3-word bundles are also examined. Again, a list of 3-word lexical bundles is extracted, focusing on key bundles in MCSAW with at least one occurrence in the reference corpus, LOCNESS (henceforth: shared 3-word bundles). In comparison with shared 4-word bundles presented earlier in Table 6.1, findings reveal that there are more shared 3-word bundles in the corpora, namely 106. However, it can be found that 43 of the 106 shared 3-word bundles are subsumed in longer strings, for instance bundle *one of the* is found to be part of the 4-word bundle *is one of the*. These bundles comprise 41% of the total shared 3-word bundles and are not investigated further, to avoid repetition. As a result, the remaining 59% (63 bundles) that are not entailed in 4-word lexical bundles are examined more closely.[64] As presented in Figure 6.1, shared 3-word bundles outnumber the shared 4-word bundles by 41%. This shows that learners produce shorter sequences of words compared to longer ones, but this is also in line with general tendencies of the English language.

[64] Following Tribble (2011: p. 99), 'entailed' here is used to indicate bundles that are comprised/included in an existing (usually longer) bundle. In addition, 'subsumed' is used interchangeably to refer to the same meaning as is used in Ädel and Erman (2012: p. 84).

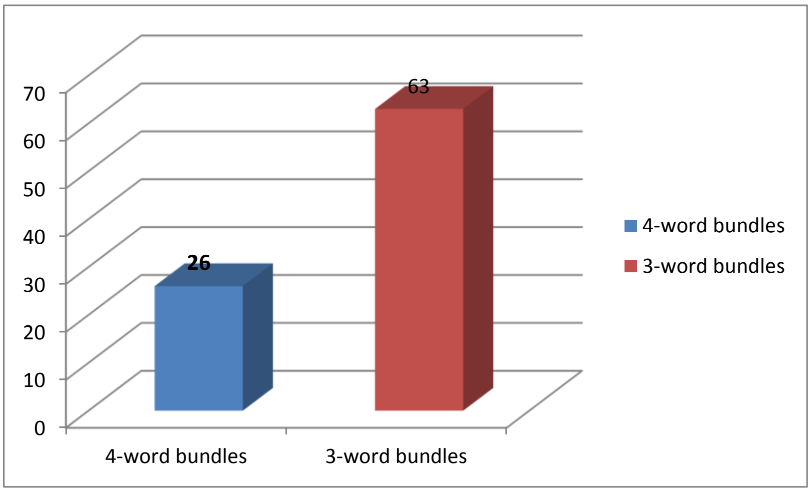


Figure 6.1: Total number of shared 3-word and 4-word bundles in MCSAW

Table 6.4: Classification of shared key 3-word lexical bundles

N	Key word	Freq.	Category	Texts	RC. Freq.	Keyness
1	in my opinion	87	Stance	82	28	41.13
2	most of the	73	Referential	56	27	29.34
3	first of all	58	Discourse Organising	58	13	37.88
4	this is because	98	Discourse Organising	73	12	90.02
5	there are some	51	Referential	47	12	32.10
6	in many ways	65	Referential	64	11	50.92
7	in front of	50	Referential	40	10	35.36
8	can be a	53	Stance	49	9	41.44
9	with each other	46	Referential	38	9	33.00
10	we have to	49	Stance	26	8	39.16
11	have their own	44	Discourse Organising	41	8	33.05
12	and so on	76	Referential	50	6	81.57
13	the chance to	49	Stance	48	6	45.01
14	it is because	37	Discourse Organising	26	6	29.68
15	you want to	35	Stance	25	5	29.97
16	that we can	109	Stance	82	4	138.25
17	to use it	63	Referential	47	4	71.68
18	the popularity of	52	Referential	51	4	56.23
19	we do not	47	Referential	43	4	49.30
20	him or her	37	Referential	35	4	35.73
21	for example if	33	Discourse Organising	28	4	30.44
22	then it is	31	Discourse Organising	29	4	27.83
23	for us to	63	Referential	54	3	76.34
24	as we know	56	Stance	46	3	66.24
25	pros and cons	37	Stance	36	3	39.39
26	face to face	34	Referential	31	3	35.26

27	to communicate with	152	Referential	99	2	215.24
28	the advantages of	83	Stance	74	2	111.35
29	and many more	59	Referential	46	2	75.73
30	of all time	39	Referential	39	2	46.56
31	friends and family	37	Referential	28	2	43.68
32	just need to	34	Stance	30	2	39.39
33	in the right	32	Stance	31	2	36.55
34	the advantage of	31	Stance	27	2	35.13
35	as a student	31	Referential	28	2	35.13
36	they are too	31	Stance	30	2	35.13
37	it is easier	30	Stance	27	2	33.72
38	to stay in	91	Referential	31	1	130.35
39	with their friends	89	Referential	67	1	127.31
40	help us to	60	Stance	44	1	83.43
41	in this world	52	Referential	44	1	71.39
42	become very important	52	Stance	52	1	71.39
43	we want to	50	Stance	38	1	68.39
44	as a conclusion	47	Discourse Organising	47	1	63.89
45	we can find	47	Stance	42	1	63.89
46	the disadvantages of	47	Stance	40	1	63.89
47	we can see	46	Stance	35	1	62.40
48	we use it	45	Referential	41	1	60.90
49	in other ways	43	Referential	42	1	57.91
50	is a social	42	Discourse Organising	39	1	56.42
51	to any other	42	Referential	41	1	56.42
52	by creating a	41	Referential	38	1	54.92
53	also can be	39	Stance	34	1	51.94
54	you can use	39	Stance	29	1	51.94
55	their time in	36	Referential	35	1	47.48
56	first and foremost	34	Discourse Organising	34	1	44.52
57	most of your	34	Referential	34	1	44.52
58	so we can	33	Stance	32	1	43.03
59	in our life	31	Referential	27	1	40.08
60	have decided to	31	Stance	31	1	40.08
61	can make us	30	Stance	26	1	38.60
62	it also can	30	Stance	26	1	38.60
63	who works in	26	Referential	26	1	32.72

Table 6.4 presents a list of the 63 shared 3-word bundles in MCSAW that occur at least once in LOCNESS. It can be seen that the bundle *in my opinion* is ranked number one, occurring 87 times in the learner corpus and 28 times in the reference language variety, while the lowest ranked bundle, occurring 26 times in MCSAW and only once in LOCNESS, is *who works in*. In addition, 12 bundles are found to be distributed in more than 50 texts, ranging from the most widely distributed bundle *to communicate with* (99 texts) to *the popularity of* (51 texts). This

means that such bundles are more recurrent across student essays than the other 51 bundles. Their recurrent occurrences across many texts compared to in the reference corpus indicate that they are more significant, thus strengthening the argument that these 12 bundles are particularly more salient in MCSAW than in LOCNESS.

Structurally, Table 6.5 shows that there are more VP-based bundles (32), followed by NP-based bundles (15) and PP-based bundles (14). In contrast to the earlier findings for shared 4-word bundles, there are more shared 3-bundles with PP fragments. Nevertheless, VP-based bundles are still the most frequent pattern, notably the prevalence of personal pronouns such as *they are too* and *help us to* (13 occurrences), the modal verb *can* as in *can make us* and *can be a* (8 occurrences), and copula *be* constructions such as *it is easier*, including the 'existential *there* + copula *be*' construction *there are some* (7 occurrences).

Table 6.5: Structural analysis of shared 3-word bundles

Lexical bundles that incorporate VP (32)	*have decided to, can make us, can be a, have their own, they are too, is a social, it also can, to use it, just need to, by creating a, also can be, it is easier, help us to, become very important, there are some, to stay in, to communicate with, we have to, we do not, we want to, you want to, we can find, we can see, we use it, you can use, who works in, this is because, it is because, so we can, that we can, as we know, then it is*
Lexical bundles that incorporate NP (15)	*the popularity of, the advantages of, the advantage of, the disadvantages of, most of your, most of the, the chance to, their time in, him or her, face to face, pros and cons, friends and family, first of all, first and foremost, and many more*
Lexical bundles that incorporate PP (14)	*in many ways, in front of, with each other, in this world, of all time, with their friends, in our life, in other ways, to any other, in my opinion, as a conclusion, as a student, in the right, for us to*
Other lexical bundles	*and so on, for example if*

Overall, there are many 3-word bundles with pronouns (i.e. *in my opinion*), 9 bundles with *we*, 3 bundles with *you/your*, 3 bundles with *us/our*, 4 bundles with *they/their, him or her*), and a variety of bundles incorporating the modal verb *can* (e.g. *can be a*, *can make us*, and *also can* *be*). The combination of both pronouns and modal verb *can* in a single bundle is also identifiable, such as *that we can*, *we can find*, *we can see*, and *so we can*. The frequent occurrence of personal pronoun *we* and modal verb *can* in shared 3-word bundles can be related to the analyses of modality in Chapter 4 and writer/reader visibility in Chapter 5, that Malaysian learners over-use *can* in expressing ability and *we* in achieving solidarity with readers of text.

The categorisation scheme developed for the analysis of 4-word bundles is also applied to the shared 3-word bundles, as shown in Table 6.4. Overall, shared 3-word bundles comprise 28 Referential bundles, 26 Stance bundles, and 9 Discourse-organising bundles. Figure 6.2 shows the graphic distribution of 3-word bundles in comparison to 4-word bundles, according to the three main functional categories, i.e. Referential bundles, Discourse-organisers, and Stance bundles. As a whole, more Referential bundles are identified, followed by Stance and Discourse-Organising bundles. It can also be seen in Figure 6.2 that shared 3-word referential bundles are produced twice as often as their respective 4-word bundles. Furthermore, discourse-organisers are found among shared 3-word bundles, which were not evident among shared 4-word bundles. These findings thus, reveal that not only do learners produce more shared 3-word bundles than shared 4-word bundles, but that they are more varied in terms of discourse functions. This, in turn, suggests that exploring more than one particular bundle length provides different results, and therefore is essential for this type of study.

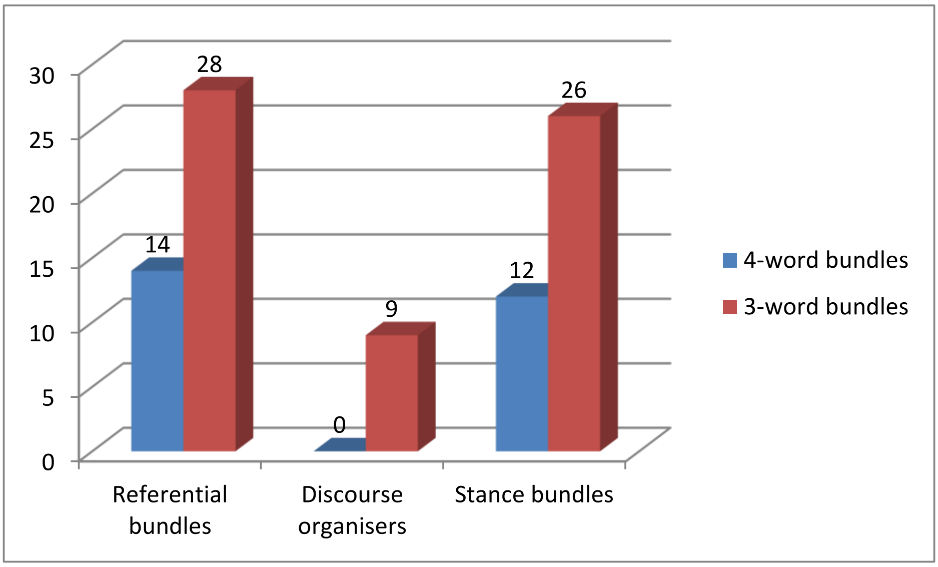


Figure 6.2: Types of 3-word and 4-word bundles in MCSAW

3-word referential bundles can further be sub-categorised, following Table 6.2, namely identification bundles (*there are some*), bundles specifying attributes (*the popularity of*, *most of the, most of your*), bundles referring to time or place (*in front of*, *in this world*, *of all time*, *in our life*), and imprecision bundles (*and so on*, *and many more, in many ways, in other ways*). The remaining 16 referential bundles are classified as 'other referential bundles', and are found to be topic related (e.g. *to communicate with*), and several paired with conjunctions (e.g. *him or her, friends and family*). In addition, adverbials (*as a student*) and negated forms (*we do not*) also fall within this sub-category.

In contrast to zero discourse organisers for 4-word shared bundles, there are 9 shared 3-word bundles that refer to the orientation of the text. These are further sub-categorised into three types: topic introduction/focus (*first of all*, *first and foremost*), topic elaboration/clarification (*is a social*, *have their own*, *for example if*), and inferential (*as a conclusion*, *it is because, this is because*). While Chen and Baker (2014, 2010) have identified high frequency use of discourse organisers in novice writing, the eleven mentioned here are low compared to referential and stance bundles shared in MCSAW and LOCNESS. Again, as previously discussed, difference in categorising bundles as Stance or Discourse organisers may have contributed to such findings. Another reason could be that 3-word discourse organisers are not heavily over-used by the Malaysian learners in MCSAW.

Finally, compared to 4-word epistemic stance bundles, there are more 3-word epistemic stance bundles identified (*in my opinion*, *as we know*). More specifically, these bundles include use of personal pronouns *my* and *we*, which have been discussed in the previous chapter as contributing to the increase of writer-visibility in learner writing. Meanwhile, use of these bundles indicates the rhetorical function of expressing personal opinion that has been argued by past researchers such as Gilquin and Paquot (2008: p. 48) as over-used by a majority of the learners in ICLE, as compared to native writers; and are all more common in the spoken component of the BNC than in the academic component (Gilquin & Paquot, 2008: p. 55).

Attitudinal stance bundles are further divided into six sub-categories. Desire bundles include *we want to, have decided to,* and *you want to,* which express a sense of personal choice with words such as *want*, *like* and *decide*. On the other hand, bundles that incorporate words such as *have to*, *should* and *need to* are grouped under bundles expressing obligation or directives (*we have to*, *just need to*). Ability bundles are found to be the most frequent type of attitudinal stance bundles (15 bundles), which consist of bundles incorporating the word *can*, and bundles incorporating words referring to *opportunity*, *chance* and *help*. These include *you can use*, *that we can*, *we can find*, *also can be*, *can make us*, *can be a*, *it also can*, *the chance to*, *help us to*, *we can see,* and *so we can*. Bundles that express importance (*become very important*) and emotivity (*the advantages of*, *the advantage of*, *the disadvantages of*, *pros and cons*, *they are too*, *it is easier,* and *in the right*) also occur. One explanation for the overuse of attitudinal stance bundles is that they indicate learners' tendency to be expressive in voicing personal opinion (attitude), more often through the personal pronoun *we*. Furthermore, as mentioned earlier, learners may also use these bundles in relation to the topic/prompt and genre of argumentative writing, which requires learners to persuade their readers to agree to their proposition, that is, advantages/disadvantages of Facebook or living in a hostel.

6.5 Qualitative analysis for key 4-word and 3-word lexical bundles

The second part of this chapter involves a more qualitative inspection, in which concordance lines of the shared 4-word and 3-word bundles are further examined. In many cases, bundles may have more than one possible meaning, and therefore it is imperative to further inspect their use in context via concordancing. In so doing, the qualitative analysis begins with a discussion on Referential bundles in Section 6.5.1, followed by Discourse-organising bundles (Section 6.5.2) and Stance bundles (Section 6.5.3). For the purpose of this section, however, only the most recurrent bundles[65] will be closely examined according to their concordance lines: these are shared 4-word and 3-word bundles that occur in more than 50 essays in MCSAW. Importantly, the analysis will consider how both bundle types are used in MCSAW, *and* compare how they are used in LOCNESS.

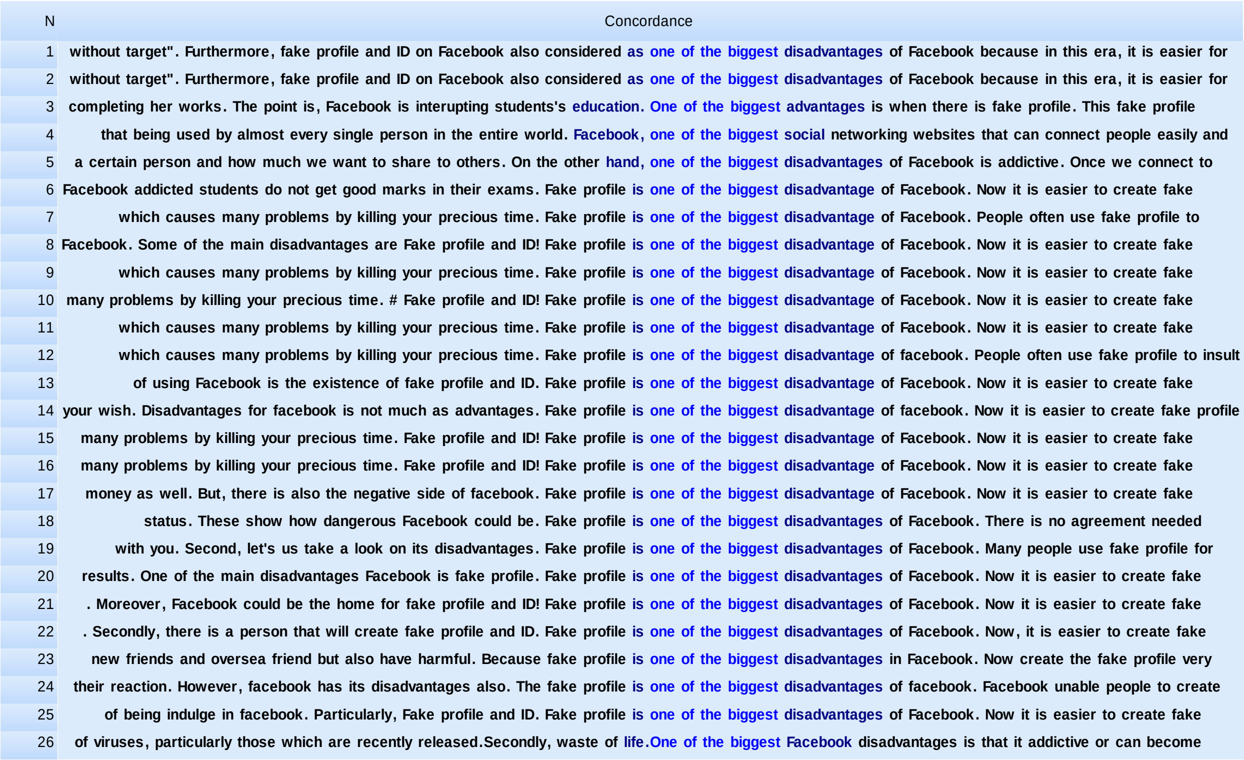
6.5.1 Referential bundles

As mentioned earlier, most shared 4-word and 3-word bundles consist of referential-type bundles. The following sub-sections will zoom in on the recurrent bundles vis-à-vis their specific sub-categories. See Table A6.1 in the Appendix for the categorisation of 4-word and 3-word referential bundles.

[65] I will use this as a cover-term for all shared 4-word and 3-word bundles that are distributed in more than 50 texts in MCSAW.

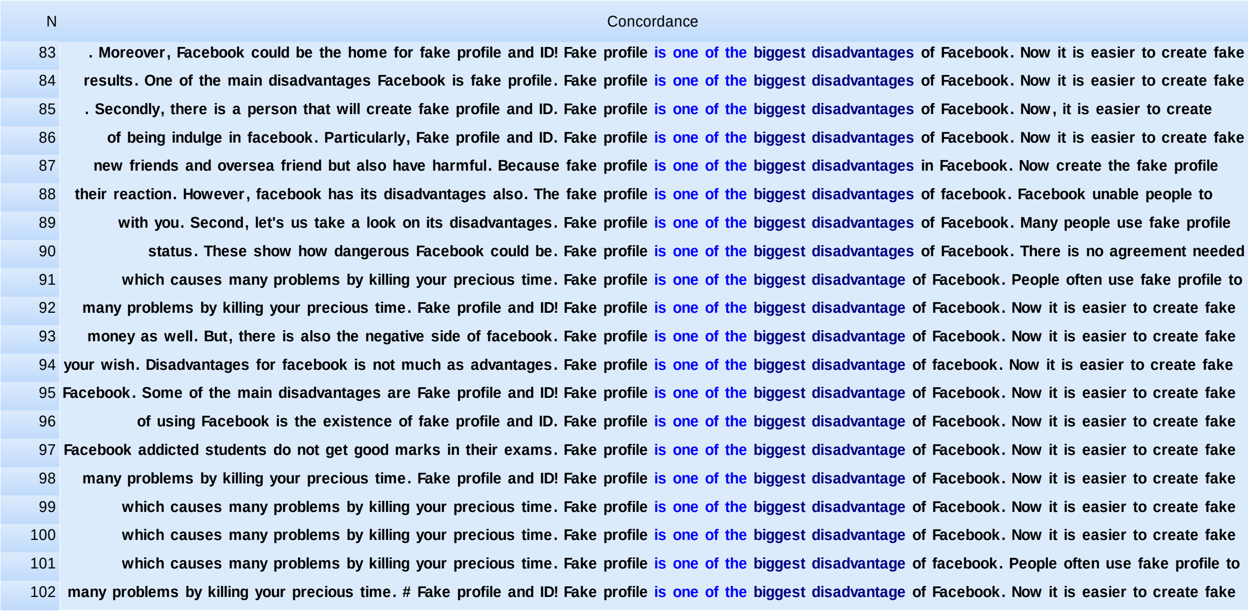
6.5.1.1 Identification/focus bundles

One of the most recurring referential-type bundles is sub-categorised as identification/focus bundles, *is one of the* (104). In order to identify whether the two 4-word-bundles *is one of the* and *one of the biggest* make up the same bundle, concordance lines for *one of the biggest* are firstly examined (Figure 6.3). It is found that 20 bundles (lines 6 – 25) of *is one of the* are entailed in the bundle *one of the biggest*. This means that they are part of a longer 5-word bundle, *is one of the biggest*. Further inspection of these 20 occurrences, as presented in Figure 6.4, shows that instances are all made up of a similar sentence, which is *Fake profile* ***is one of the biggest*** *disadvantage*(*s*) *of Facebook*. This bundle, therefore, is mainly over-used in reference to the disadvantages of Facebook. Further inspection also reveals that the lines originated from different text files, so that it is possible that the identical concordance lines are suggestive of copying on the part of the learners or influenced by pedagogical effects that have been taught during practices in the classroom. This reflects tautological findings from previous chapters.



N	Concordance
1	without target". Furthermore, fake profile and ID on Facebook also considered as one of the biggest disadvantages of Facebook because in this era, it is easier for
2	without target". Furthermore, fake profile and ID on Facebook also considered as one of the biggest disadvantages of Facebook because in this era, it is easier for
3	completing her works. The point is, Facebook is interupting students's education. One of the biggest advantages is when there is fake profile. This fake profile
4	that being used by almost every single person in the entire world. Facebook, one of the biggest social networking websites that can connect people easily and
5	a certain person and how much we want to share to others. On the other hand, one of the biggest disadvantages of Facebook is addictive. Once we connect to
6	Facebook addicted students do not get good marks in their exams. Fake profile is one of the biggest disadvantage of Facebook. Now it is easier to create fake
7	which causes many problems by killing your precious time. Fake profile is one of the biggest disadvantage of Facebook. People often use fake profile to
8	Facebook. Some of the main disadvantages are Fake profile and ID! Fake profile is one of the biggest disadvantage of Facebook. Now it is easier to create fake
9	which causes many problems by killing your precious time. Fake profile is one of the biggest disadvantage of Facebook. Now it is easier to create fake
10	many problems by killing your precious time. # Fake profile and ID! Fake profile is one of the biggest disadvantage of Facebook. Now it is easier to create fake
11	which causes many problems by killing your precious time. Fake profile is one of the biggest disadvantage of Facebook. Now it is easier to create fake
12	which causes many problems by killing your precious time. Fake profile is one of the biggest disadvantage of facebook. People often use fake profile to insult
13	of using Facebook is the existence of fake profile and ID. Fake profile is one of the biggest disadvantage of Facebook. Now it is easier to create fake
14	your wish. Disadvantages for facebook is not much as advantages. Fake profile is one of the biggest disadvantage of facebook. Now it is easier to create fake profile
15	many problems by killing your precious time. Fake profile and ID! Fake profile is one of the biggest disadvantage of Facebook. Now it is easier to create fake
16	many problems by killing your precious time. Fake profile and ID! Fake profile is one of the biggest disadvantage of Facebook. Now it is easier to create fake
17	money as well. But, there is also the negative side of facebook. Fake profile is one of the biggest disadvantage of Facebook. Now it is easier to create fake
18	status. These show how dangerous Facebook could be. Fake profile is one of the biggest disadvantages of Facebook. There is no agreement needed
19	with you. Second, let's us take a look on its disadvantages. Fake profile is one of the biggest disadvantages of Facebook. Many people use fake profile for
20	results. One of the main disadvantages Facebook is fake profile. Fake profile is one of the biggest disadvantages of Facebook. Now it is easier to create fake
21	. Moreover, Facebook could be the home for fake profile and ID! Fake profile is one of the biggest disadvantages of Facebook. Now it is easier to create fake
22	. Secondly, there is a person that will create fake profile and ID. Fake profile is one of the biggest disadvantages of Facebook. Now, it is easier to create fake
23	new friends and oversea friend but also have harmful. Because fake profile is one of the biggest disadvantages in Facebook. Now create the fake profile very
24	their reaction. However, facebook has its disadvantages also. The fake profile is one of the biggest disadvantages of facebook. Facebook unable people to create
25	of being indulge in facebook. Particularly, Fake profile and ID. Fake profile is one of the biggest disadvantages of Facebook. Now it is easier to create fake
26	of viruses, particularly those which are recently released.Secondly, waste of life.One of the biggest Facebook disadvantages is that it addictive or can become

Figure 6.3: Concordance lines for *one of the biggest* in MCSAW

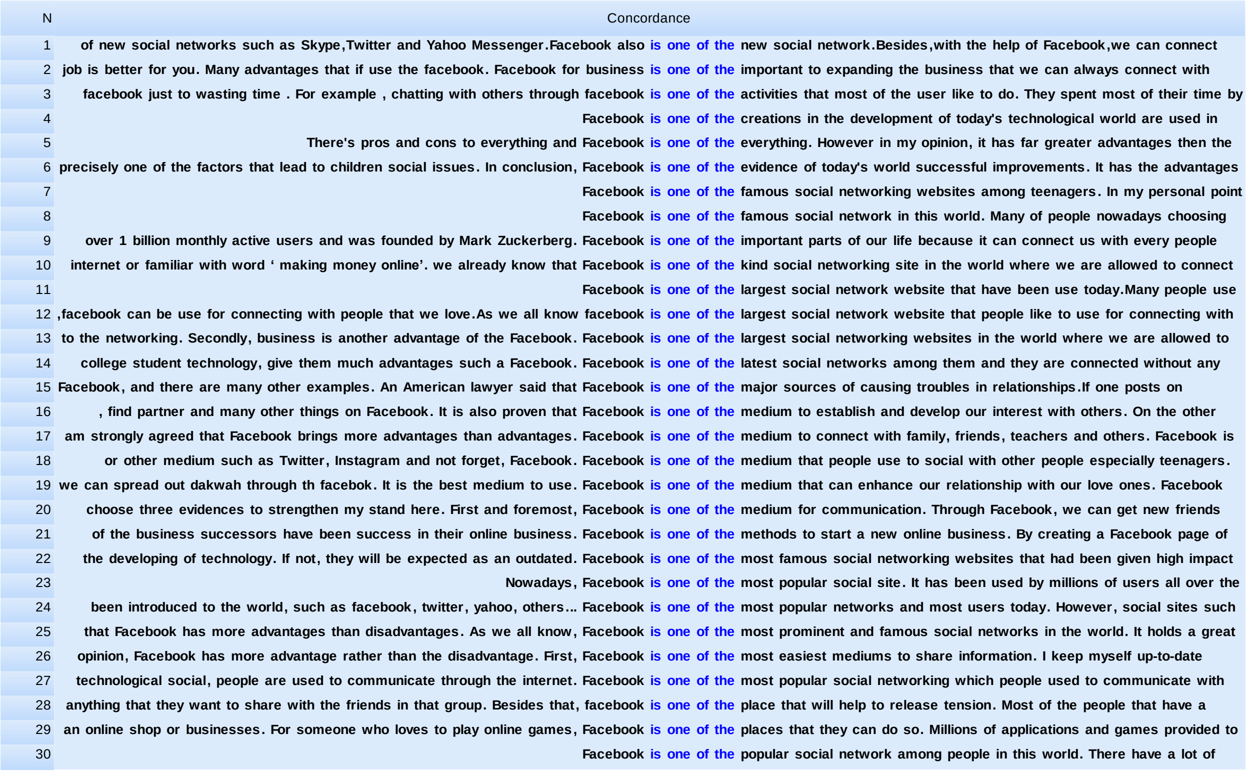


N	Concordance
83	. Moreover, Facebook could be the home for fake profile and ID! Fake profile is one of the biggest disadvantages of Facebook. Now it is easier to create fake
84	results. One of the main disadvantages Facebook is fake profile. Fake profile is one of the biggest disadvantages of Facebook. Now it is easier to create fake
85	. Secondly, there is a person that will create fake profile and ID. Fake profile is one of the biggest disadvantages of Facebook. Now, it is easier to create
86	of being indulge in facebook. Particularly, Fake profile and ID. Fake profile is one of the biggest disadvantages of Facebook. Now it is easier to create fake
87	new friends and oversea friend but also have harmful. Because fake profile is one of the biggest disadvantages in Facebook. Now create the fake profile
88	their reaction. However, facebook has its disadvantages also. The fake profile is one of the biggest disadvantages of facebook. Facebook unable people to
89	with you. Second, let's us take a look on its disadvantages. Fake profile is one of the biggest disadvantages of Facebook. Many people use fake profile
90	status. These show how dangerous Facebook could be. Fake profile is one of the biggest disadvantages of Facebook. There is no agreement needed
91	which causes many problems by killing your precious time. Fake profile is one of the biggest disadvantage of Facebook. People often use fake profile to
92	many problems by killing your precious time. Fake profile and ID! Fake profile is one of the biggest disadvantage of Facebook. Now it is easier to create fake
93	money as well. But, there is also the negative side of facebook. Fake profile is one of the biggest disadvantage of Facebook. Now it is easier to create fake
94	your wish. Disadvantages for facebook is not much as advantages. Fake profile is one of the biggest disadvantage of facebook. Now it is easier to create fake
95	Facebook. Some of the main disadvantages are Fake profile and ID! Fake profile is one of the biggest disadvantage of Facebook. Now it is easier to create fake
96	of using Facebook is the existence of fake profile and ID. Fake profile is one of the biggest disadvantage of Facebook. Now it is easier to create fake
97	Facebook addicted students do not get good marks in their exams. Fake profile is one of the biggest disadvantage of Facebook. Now it is easier to create fake
98	many problems by killing your precious time. Fake profile and ID! Fake profile is one of the biggest disadvantage of Facebook. Now it is easier to create fake
99	which causes many problems by killing your precious time. Fake profile is one of the biggest disadvantage of Facebook. Now it is easier to create fake
100	which causes many problems by killing your precious time. Fake profile is one of the biggest disadvantage of Facebook. Now it is easier to create fake
101	which causes many problems by killing your precious time. Fake profile is one of the biggest disadvantage of facebook. People often use fake profile to
102	many problems by killing your precious time. # Fake profile and ID! Fake profile is one of the biggest disadvantage of Facebook. Now it is easier to create fake

Figure 6.4: Bundles *is one of the* that constitute bundles *one of the biggest*

Figure 6.5 presents 30 instances of how bundle *is one of the* is used in MCSAW. This structure is usually made up of an NP + copula *be* + NP/AdjP. Apart from *biggest*, adjectives following this bundle include *new*, *important*, *famous*, *largest*, *latest*, *major* and *popular*. Other adjectives can be found with the superlative *most*, such as *most famous*/*popular*/*prominent*/*easiest**.[66] While not incorporating a stance form itself, this bundle is thus, strongly associated with the expression of stance in its right-hand co-text. Most of the head nouns that are given focus by this bundle are largely associated with the topic Facebook (e.g. *social network*/*social networking sites*/*websites*, *activities*, *creations*, *evidence*, *sources*, *medium*, *method*, *place* etc.), while one instance was related to the hostel topic, *Hostel* ***is one of the*** *safe places where most of the college students stay at since they live far away from their home* (MCSAW_288.txt). A common error is also found: learners use a singular noun following the phrase *is one of the* as opposed to the correct plural form (e.g. *Facebook is one of the way* to share our opinion*).

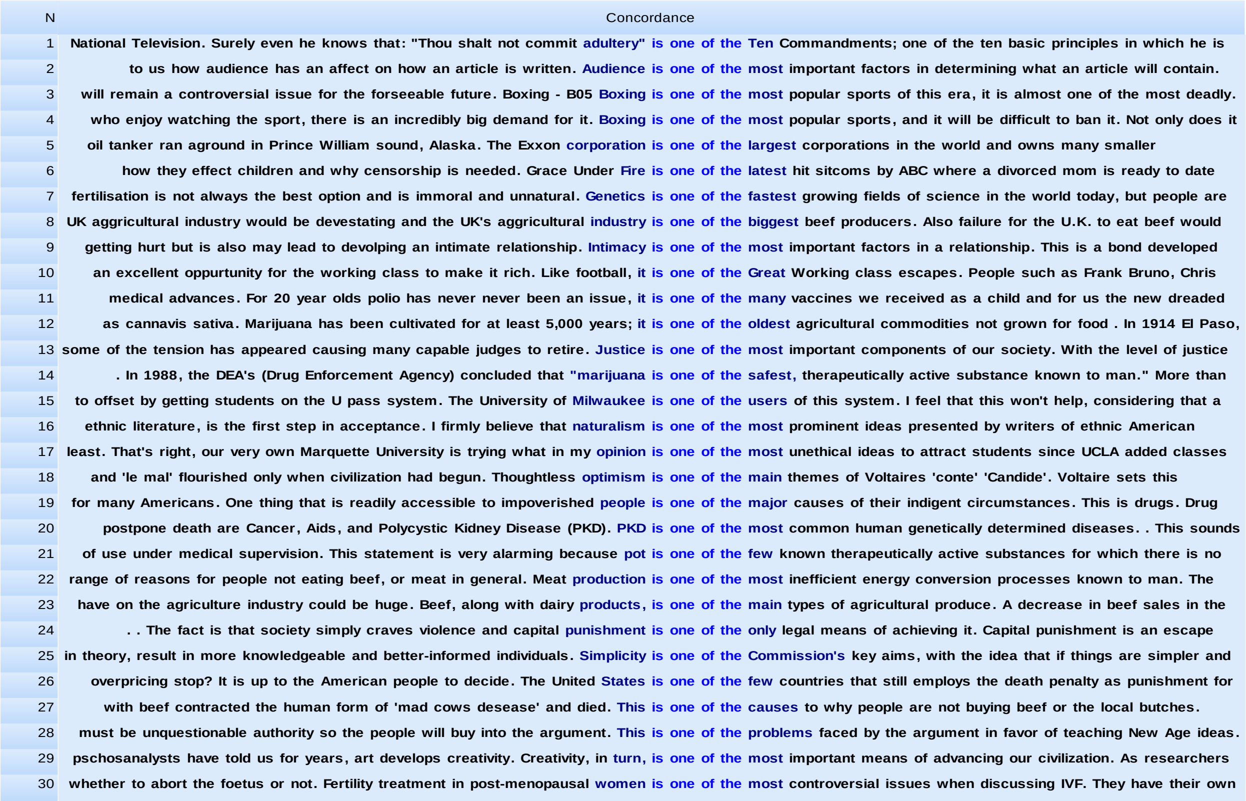
[66] The * symbol here means incorrect grammar/ grammar error given the instance referred to from the concordance line (*most easiest*).



N	Concordance
1	of new social networks such as Skype,Twitter and Yahoo Messenger.Facebook also is one of the new social network.Besides,with the help of Facebook,we can connect
2	job is better for you. Many advantages that if use the facebook. Facebook for business is one of the important to expanding the business that we can always connect with
3	facebook just to wasting time . For example , chatting with others through facebook is one of the activities that most of the user like to do. They spent most of their time by
4	Facebook is one of the creations in the development of today's technological world are used in
5	There's pros and cons to everything and Facebook is one of the everything. However in my opinion, it has far greater advantages then the
6	precisely one of the factors that lead to children social issues. In conclusion, Facebook is one of the evidence of today's world successful improvements. It has the advantages
7	Facebook is one of the famous social networking websites among teenagers. In my personal point
8	Facebook is one of the famous social network in this world. Many of people nowadays choosing
9	over 1 billion monthly active users and was founded by Mark Zuckerberg. Facebook is one of the important parts of our life because it can connect us with every people
10	internet or familiar with word ‘ making money online’. we already know that Facebook is one of the kind social networking site in the world where we are allowed to connect
11	Facebook is one of the largest social network website that have been use today.Many people use
12	,facebook can be use for connecting with people that we love.As we all know facebook is one of the largest social network website that people like to use for connecting with
13	to the networking. Secondly, business is another advantage of the Facebook. Facebook is one of the largest social networking websites in the world where we are allowed to
14	college student technology, give them much advantages such a Facebook. Facebook is one of the latest social networks among them and they are connected without any
15	Facebook, and there are many other examples. An American lawyer said that Facebook is one of the major sources of causing troubles in relationships.If one posts on
16	, find partner and many other things on Facebook. It is also proven that Facebook is one of the medium to establish and develop our interest with others. On the other
17	am strongly agreed that Facebook brings more advantages than advantages. Facebook is one of the medium to connect with family, friends, teachers and others. Facebook is
18	or other medium such as Twitter, Instagram and not forget, Facebook. Facebook is one of the medium that people use to social with other people especially teenagers.
19	we can spread out dakwah through th facebok. It is the best medium to use. Facebook is one of the medium that can enhance our relationship with our love ones. Facebook
20	choose three evidences to strengthen my stand here. First and foremost, Facebook is one of the medium for communication. Through Facebook, we can get new friends
21	of the business successors have been success in their online business. Facebook is one of the methods to start a new online business. By creating a Facebook page of
22	the developing of technology. If not, they will be expected as an outdated. Facebook is one of the most famous social networking websites that had been given high impact
23	Nowadays, Facebook is one of the most popular social site. It has been used by millions of users all over the
24	been introduced to the world, such as facebook, twitter, yahoo, others... Facebook is one of the most popular networks and most users today. However, social sites such
25	that Facebook has more advantages than disadvantages. As we all know, Facebook is one of the most prominent and famous social networks in the world. It holds a great
26	opinion, Facebook has more advantage rather than the disadvantage. First, Facebook is one of the most easiest mediums to share information. I keep myself up-to-date
27	technological social, people are used to communicate through the internet. Facebook is one of the most popular social networking which people used to communicate with
28	anything that they want to share with the friends in that group. Besides that, facebook is one of the place that will help to release tension. Most of the people that have a
29	an online shop or businesses. For someone who loves to play online games, Facebook is one of the places that they can do so. Millions of applications and games provided to
30	Facebook is one of the popular social network among people in this world. There have a lot of

Figure 6.5: Concordance lines for *is one of the* in MCSAW

It is also interesting to note that this bundle is used in the same way in LOCNESS (as shown in Figure 6.6), occurring 30 times. However, despite the similarities, *biggest* was not over-used in relation to the construction of the verb phrase in LOCNESS (only 3 times out of 30 occurrences). Although the bundle in both corpora is seen to be used in relation to the topic, learners seem to over-use this bundle in association with the word *disadvantage*(*s*), and hence this shows that *is one of the biggest* is more prevalent in Malaysian learners' writing. As argued in Lee and Chen (2009: p. 160), learners tend to over-use simple, common words such as the word form *big* in *biggest*, at the expense of fine-tuning their style and rhetorical manner of writing. Furthermore, the use of *is one of the biggest* prevents writers from being accused of over-generalising in their statements, in contrast to using only the superlative ‘the biggest’ in making their claims. It is also worth noting that the bundle *is one of the* can be traced to its Malay translation *merupakan salah satu daripada*, which is commonly found in formal spoken and written Malay discourse. The use of this bundle in the first language (or L1) is also seen to be similar in syntax and portrayal of the discourse function, thus indicating that direct translation might have occurred.



N	Concordance
1	National Television. Surely even he knows that: "Thou shalt not commit adultery" is one of the Ten Commandments; one of the ten basic principles in which he is
2	to us how audience has an affect on how an article is written. Audience is one of the most important factors in determining what an article will contain.
3	will remain a controversial issue for the forseeable future. Boxing - B05 Boxing is one of the most popular sports of this era, it is almost one of the most deadly.
4	who enjoy watching the sport, there is an incredibly big demand for it. Boxing is one of the most popular sports, and it will be difficult to ban it. Not only does it
5	oil tanker ran aground in Prince William sound, Alaska. The Exxon corporation is one of the largest corporations in the world and owns many smaller
6	how they effect children and why censorship is needed. Grace Under Fire is one of the latest hit sitcoms by ABC where a divorced mom is ready to date
7	fertilisation is not always the best option and is immoral and unnatural. Genetics is one of the fastest growing fields of science in the world today, but people are
8	UK aggricultural industry would be devestating and the UK's aggricultural industry is one of the biggest beef producers. Also failure for the U.K. to eat beef would
9	getting hurt but is also may lead to devolping an intimate relationship. Intimacy is one of the most important factors in a relationship. This is a bond developed
10	an excellent oppurtunity for the working class to make it rich. Like football, it is one of the Great Working class escapes. People such as Frank Bruno, Chris
11	medical advances. For 20 year olds polio has never never been an issue, it is one of the many vaccines we received as a child and for us the new dreaded
12	as cannavis sativa. Marijuana has been cultivated for at least 5,000 years; it is one of the oldest agricultural commodities not grown for food . In 1914 El Paso,
13	some of the tension has appeared causing many capable judges to retire. Justice is one of the most important components of our society. With the level of justice
14	. In 1988, the DEA's (Drug Enforcement Agency) concluded that "marijuana is one of the safest, therapeutically active substance known to man." More than
15	to offset by getting students on the U pass system. The University of Milwaukee is one of the users of this system. I feel that this won't help, considering that a
16	ethnic literature, is the first step in acceptance. I firmly believe that naturalism is one of the most prominent ideas presented by writers of ethnic American
17	least. That's right, our very own Marquette University is trying what in my opinion is one of the most unethical ideas to attract students since UCLA added classes
18	and 'le mal' flourished only when civilization had begun. Thoughtless optimism is one of the main themes of Voltaires 'conte' 'Candide'. Voltaire sets this
19	for many Americans. One thing that is readily accessible to impoverished people is one of the major causes of their indigent circumstances. This is drugs. Drug
20	postpone death are Cancer, Aids, and Polycystic Kidney Disease (PKD). PKD is one of the most common human genetically determined diseases. . This sounds
21	of use under medical supervision. This statement is very alarming because pot is one of the few known therapeutically active substances for which there is no
22	range of reasons for people not eating beef, or meat in general. Meat production is one of the most inefficient energy conversion processes known to man. The
23	have on the agriculture industry could be huge. Beef, along with dairy products, is one of the main types of agricultural produce. A decrease in beef sales in the
24	. . The fact is that society simply craves violence and capital punishment is one of the only legal means of achieving it. Capital punishment is an escape
25	in theory, result in more knowledgeable and better-informed individuals. Simplicity is one of the Commission's key aims, with the idea that if things are simpler and
26	overpricing stop? It is up to the American people to decide. The United States is one of the few countries that still employs the death penalty as punishment for
27	with beef contracted the human form of 'mad cows desease' and died. This is one of the causes to why people are not buying beef or the local butches.
28	must be unquestionable authority so the people will buy into the argument. This is one of the problems faced by the argument in favor of teaching New Age ideas.
29	pschosanalysts have told us for years, art develops creativity. Creativity, in turn, is one of the most important means of advancing our civilization. As researchers
30	whether to abort the foetus or not. Fertility treatment in post-menopausal women is one of the most controversial issues when discussing IVF. They have their own

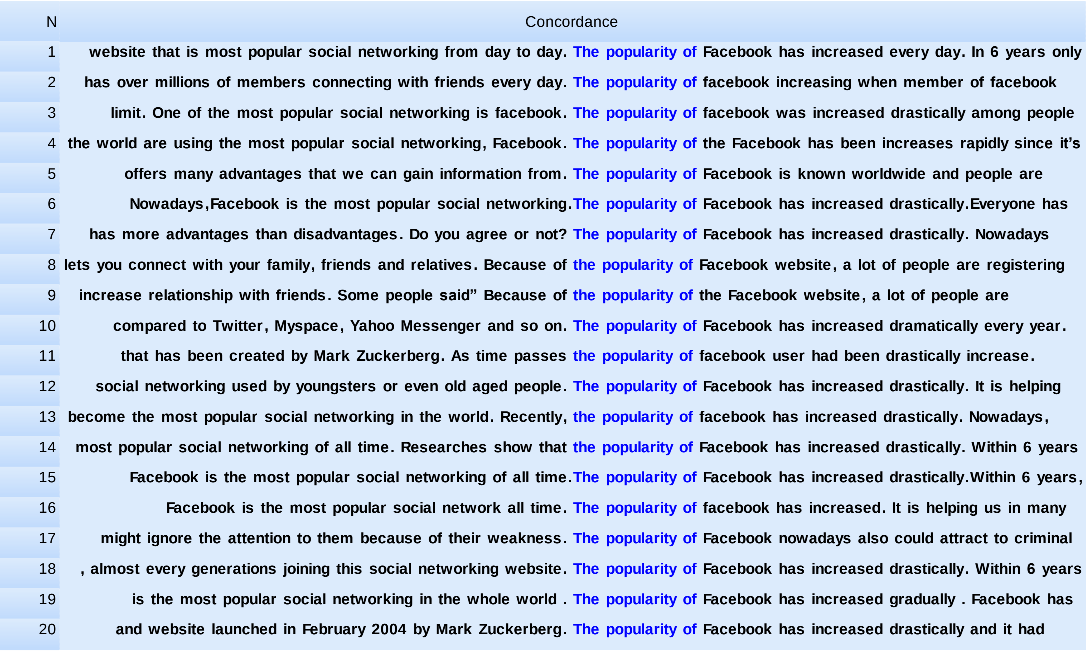
Figure 6.6: Concordance lines for *is one of the* in LOCNESS

6.5.1.2 Bundles specifying attributes

Recurrent referential bundles that specify attributes to identified entities are sub-divided into two types: bundles specifying attributes of following entities (*the popularity of*, *most of the*), and bundles specifying attributes of preceding entities (*people around the world*). These bundles are further discussed in the following sub-sections.

6.5.1.2.1 Bundles specifying attributes of following nouns/entities

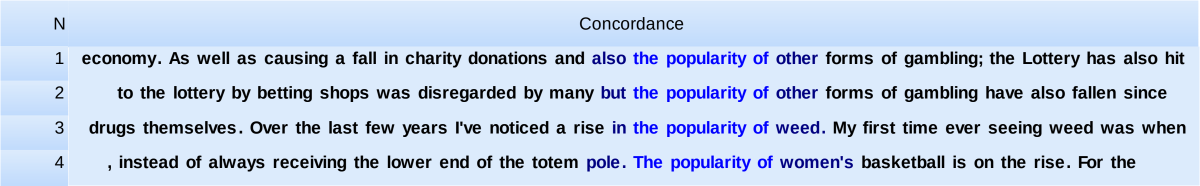
Bundles *the popularity of* and *most of the* are categorised as specifying attributes of following entities, which are not incorporated within the bundles themselves. Out of the total 51 times that *the popularity of* occur in MCSAW, 36 examples are used in the same way: *the popularity of Facebook has increased*. Essentially, the bundle is used in this way to suggest that Facebook is popular and that it is continuously becoming more so. The 20 lines presented in Figure 6.7 show that the bundle co-occurs with emphasisers, most often the adverb *drastically* (10 times), pertaining to the VP *has increased*. Interestingly, learners use this intensifier, including *dramatically* (line 10), 41 times as regards the clause *the popularity of Facebook has increased*. One possible explanation could be due to teaching aids that might have been used repeatedly in the classroom, and as a result learners have familiarised themselves by pairing the verb *increase* with *drastically* or *dramatically*. Other adverbs pertaining to the verb *increase* include *rapidly* (line 4) and *gradually* (line 19).



N	Concordance
1	website that is most popular social networking from day to day. The popularity of Facebook has increased every day. In 6 years only
2	has over millions of members connecting with friends every day. The popularity of facebook increasing when member of facebook
3	limit. One of the most popular social networking is facebook. The popularity of facebook was increased drastically among people
4	the world are using the most popular social networking, Facebook. The popularity of the Facebook has been increases rapidly since it's
5	offers many advantages that we can gain information from. The popularity of Facebook is known worldwide and people are
6	Nowadays,Facebook is the most popular social networking.The popularity of Facebook has increased drastically.Everyone has
7	has more advantages than disadvantages. Do you agree or not? The popularity of Facebook has increased drastically. Nowadays
8	lets you connect with your family, friends and relatives. Because of the popularity of Facebook website, a lot of people are registering
9	increase relationship with friends. Some people said" Because of the popularity of the Facebook website, a lot of people are
10	compared to Twitter, Myspace, Yahoo Messenger and so on. The popularity of Facebook has increased dramatically every year.
11	that has been created by Mark Zuckerberg. As time passes the popularity of facebook user had been drastically increase.
12	social networking used by youngsters or even old aged people. The popularity of Facebook has increased drastically. It is helping
13	become the most popular social networking in the world. Recently, the popularity of facebook has increased drastically. Nowadays,
14	most popular social networking of all time. Researches show that the popularity of Facebook has increased drastically. Within 6 years
15	Facebook is the most popular social networking of all time.The popularity of Facebook has increased drastically.Within 6 years,
16	Facebook is the most popular social network all time. The popularity of facebook has increased. It is helping us in many
17	might ignore the attention to them because of their weakness. The popularity of Facebook nowadays also could attract to criminal
18	, almost every generations joining this social networking website. The popularity of Facebook has increased drastically. Within 6 years
19	is the most popular social networking in the whole world . The popularity of Facebook has increased gradually . Facebook has
20	and website launched in February 2004 by Mark Zuckerberg. The popularity of Facebook has increased drastically and it had

Figure 6.7: Concordance lines for *the popularity of* in MCSAW

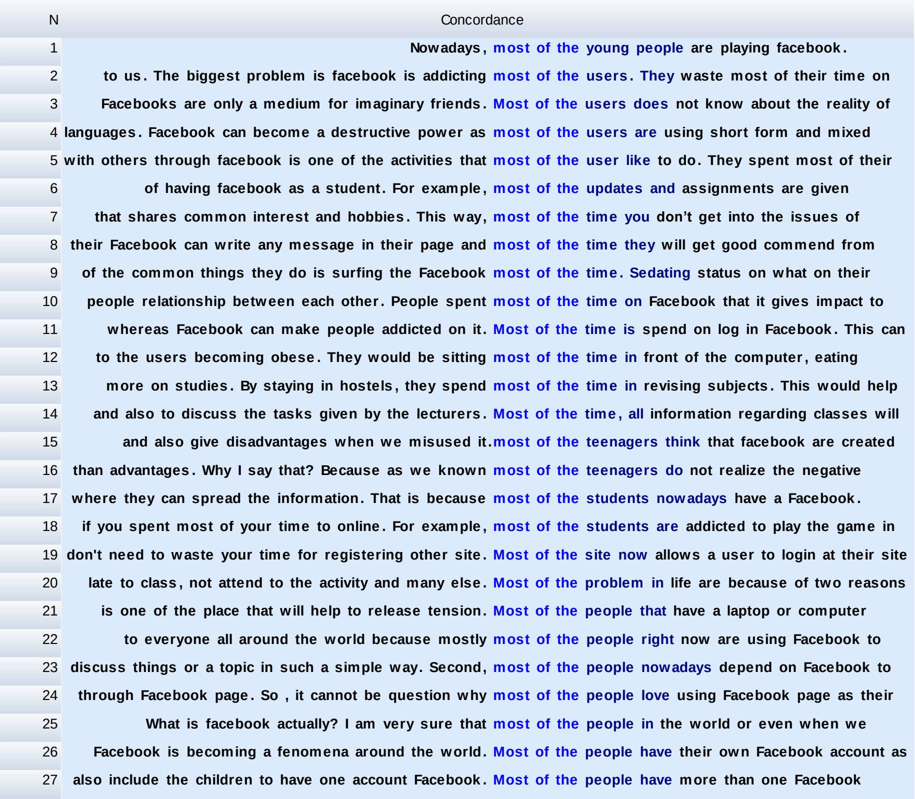
Figure 6.8, on the other hand, shows only 4 instances in which the bundle occurs in LOCNESS. Similar to its use in MCSAW, *the popularity of* in lines 3 and 4 in Figure 6.8 specify attributes (i.e. pervasiveness) of the following entities; marijuana (*weed*) and women's basketball. Interestingly, the first two lines show similar instances of the bundle in a longer string of words: ***the popularity of*** *other forms of gambling*. This indicates that the bundle is joined with another referential-type bundle, *other forms of*, which characterises the entity *gambling*. However, this was not found in learner writing.



N	Concordance
1	economy. As well as causing a fall in charity donations and also the popularity of other forms of gambling; the Lottery has also hit
2	to the lottery by betting shops was disregarded by many but the popularity of other forms of gambling have also fallen since
3	drugs themselves. Over the last few years I've noticed a rise in the popularity of weed. My first time ever seeing weed was when
4	, instead of always receiving the lower end of the totem pole. The popularity of women's basketball is on the rise. For the

Figure 6.8: Concordance lines for *the popularity of* in LOCNESS

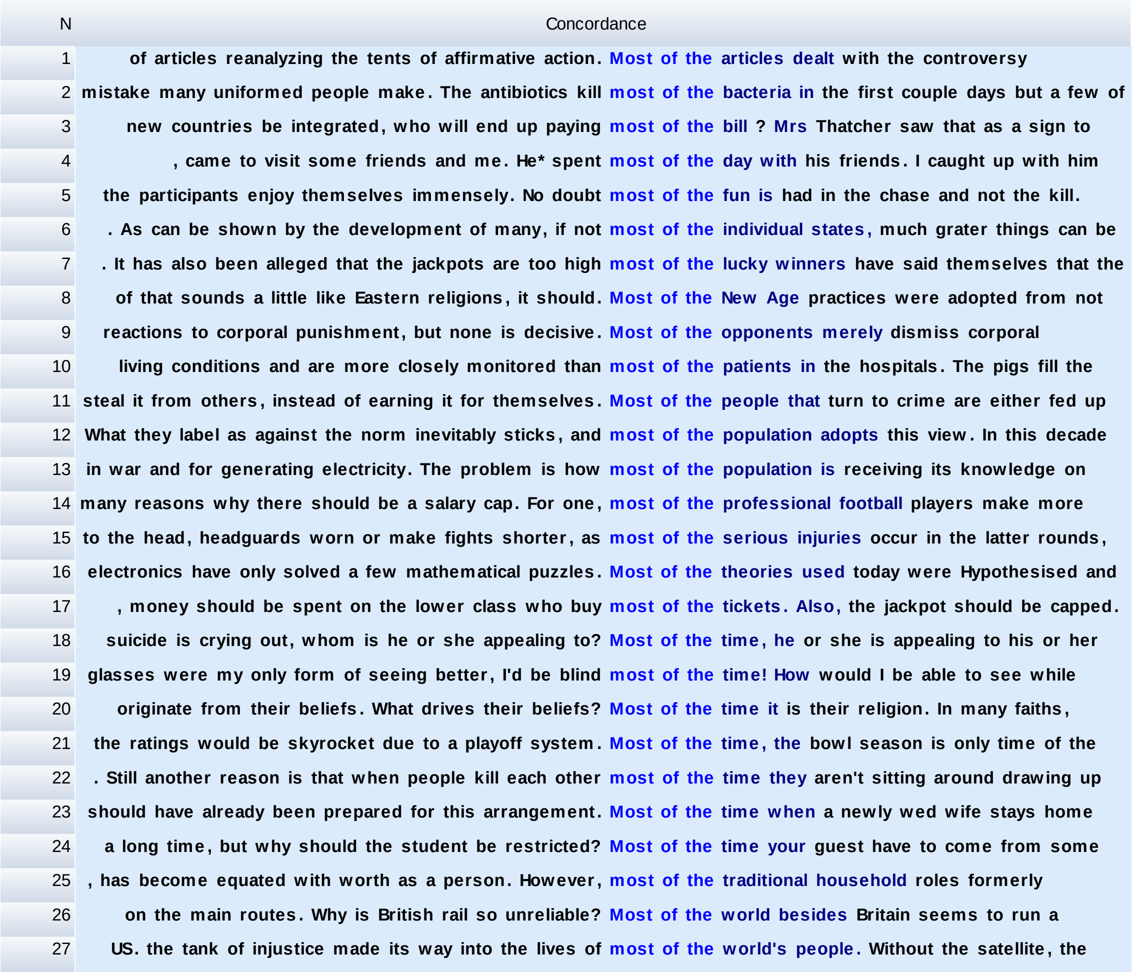
Most of the is more distributed (56 texts) than *the popularity of* (51 texts). Out of the total occurrences, *most of the* is found 11 times in one particular sentence: ***Most of the people** who know how to use a computer and internet, have a profile on Facebook*. Figure 6.9 provides further evidence that this bundle is used to refer to the number of particular categories of people. Apart from *people* (26), other head nouns include *users* (6), *teenagers* (2), *students* (7), *lecturers*, *Facebook members/users/Facebookers*, *entrepreneur* (2), *customer*, *criminals*, *children and teenagers*, *businessman*, and *business people* (2). Non-human or inanimate head nouns include *updates and assignments*, *problem*, *online shop case*, *site* (3), *Fan page*, and, most frequently, *time* (10).



N	Concordance
1	Nowadays, most of the young people are playing facebook.
2	to us. The biggest problem is facebook is addicting most of the users. They waste most of their time on
3	Facebooks are only a medium for imaginary friends. Most of the users does not know about the reality of
4	languages. Facebook can become a destructive power as most of the users are using short form and mixed
5	with others through facebook is one of the activities that most of the user like to do. They spent most of their
6	of having facebook as a student. For example, most of the updates and assignments are given
7	that shares common interest and hobbies. This way, most of the time you don't get into the issues of
8	their Facebook can write any message in their page and most of the time they will get good commend from
9	of the common things they do is surfing the Facebook most of the time. Sedating status on what on their
10	people relationship between each other. People spent most of the time on Facebook that it gives impact to
11	whereas Facebook can make people addicted on it. Most of the time is spend on log in Facebook. This can
12	to the users becoming obese. They would be sitting most of the time in front of the computer, eating
13	more on studies. By staying in hostels, they spend most of the time in revising subjects. This would help
14	and also to discuss the tasks given by the lecturers. Most of the time, all information regarding classes will
15	and also give disadvantages when we misused it.most of the teenagers think that facebook are created
16	than advantages. Why I say that? Because as we known most of the teenagers do not realize the negative
17	where they can spread the information. That is because most of the students nowadays have a Facebook.
18	if you spent most of your time to online. For example, most of the students are addicted to play the game in
19	don't need to waste your time for registering other site. Most of the site now allows a user to login at their site
20	late to class, not attend to the activity and many else. Most of the problem in life are because of two reasons
21	is one of the place that will help to release tension. Most of the people that have a laptop or computer
22	to everyone all around the world because mostly most of the people right now are using Facebook to
23	discuss things or a topic in such a simple way. Second, most of the people nowadays depend on Facebook to
24	through Facebook page. So , it cannot be question why most of the people love using Facebook page as their
25	What is facebook actually? I am very sure that most of the people in the world or even when we
26	Facebook is becoming a fenomena around the world. Most of the people have their own Facebook account as
27	also include the children to have one account Facebook. Most of the people have more than one Facebook

Figure 6.9: Concordance lines for *most of the* in MCSAW

Similarly, in LOCNESS all 27 occurrences of *most of the* (in Figure 6.10) are used to indicate quantity or amount of the following entities, namely *articles*, *bacteria*, *bill*, *day*, *fun*, *individual states*, *lucky winners*, *New Age practices*, *opponents*, *patients*, *people*, *population* (2), *professional football players*, *serious injuries*, *theories*, *tickets*, *time* (7), *traditional household roles*, *world*, and *world's people*. The lower number of 'human' head nouns and the greater variation of nouns in the reference language variety compared to the learner corpus indicate topic variability, which is limited in MCSAW. Importantly, writers in both LOCNESS and MCSAW make frequent use of the phrase *most of the time*.

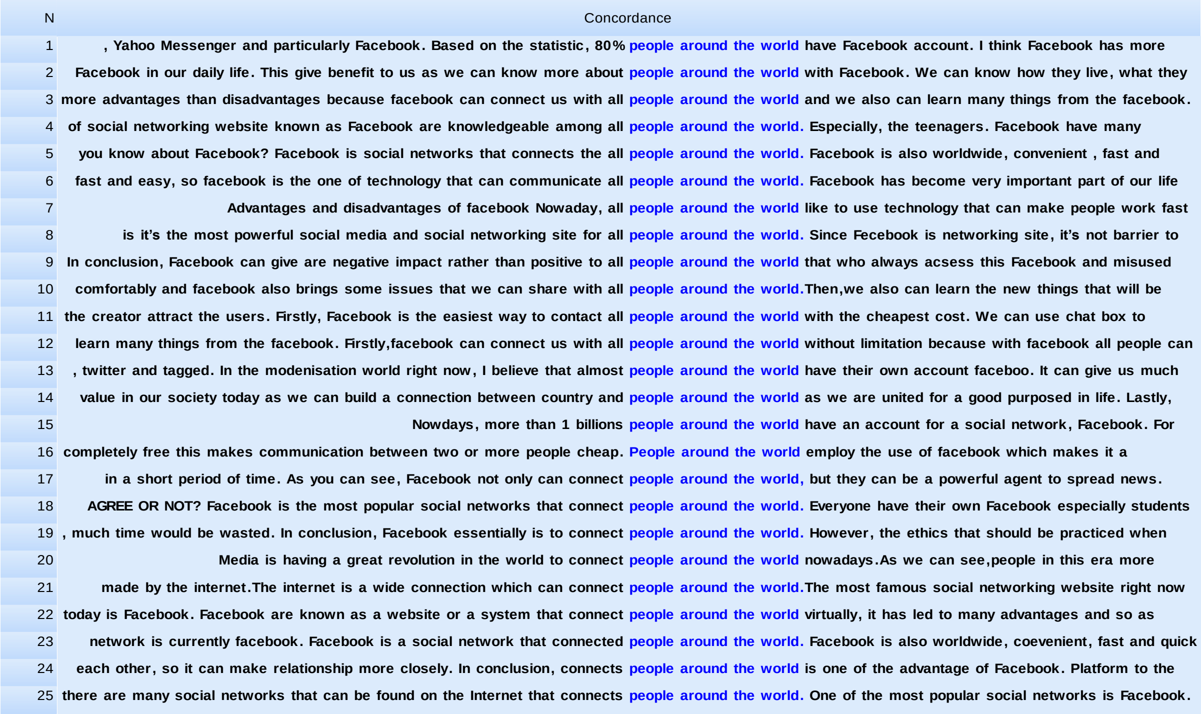


N	Concordance
1	of articles reanalyzing the tents of affirmative action. Most of the articles dealt with the controversy
2	mistake many uniformed people make. The antibiotics kill most of the bacteria in the first couple days but a few of
3	new countries be integrated, who will end up paying most of the bill ? Mrs Thatcher saw that as a sign to
4	, came to visit some friends and me. He* spent most of the day with his friends. I caught up with him
5	the participants enjoy themselves immensely. No doubt most of the fun is had in the chase and not the kill.
6	. As can be shown by the development of many, if not most of the individual states, much grater things can be
7	. It has also been alleged that the jackpots are too high most of the lucky winners have said themselves that the
8	of that sounds a little like Eastern religions, it should. Most of the New Age practices were adopted from not
9	reactions to corporal punishment, but none is decisive. Most of the opponents merely dismiss corporal
10	living conditions and are more closely monitored than most of the patients in the hospitals. The pigs fill the
11	steal it from others, instead of earning it for themselves. Most of the people that turn to crime are either fed up
12	What they label as against the norm inevitably sticks, and most of the population adopts this view. In this decade
13	in war and for generating electricity. The problem is how most of the population is receiving its knowledge on
14	many reasons why there should be a salary cap. For one, most of the professional football players make more
15	to the head, headguards worn or make fights shorter, as most of the serious injuries occur in the latter rounds,
16	electronics have only solved a few mathematical puzzles. Most of the theories used today were Hypothesised and
17	, money should be spent on the lower class who buy most of the tickets. Also, the jackpot should be capped.
18	suicide is crying out, whom is he or she appealing to? Most of the time, he or she is appealing to his or her
19	glasses were my only form of seeing better, I'd be blind most of the time! How would I be able to see while
20	originate from their beliefs. What drives their beliefs? Most of the time it is their religion. In many faiths,
21	the ratings would be skyrocket due to a playoff system. Most of the time, the bowl season is only time of the
22	. Still another reason is that when people kill each other most of the time they aren't sitting around drawing up
23	should have already been prepared for this arrangement. Most of the time when a newly wed wife stays home
24	a long time, but why should the student be restricted? Most of the time your guest have to come from some
25	, has become equated with worth as a person. However, most of the traditional household roles formerly
26	on the main routes. Why is British rail so unreliable? Most of the world besides Britain seems to run a
27	US. the tank of injustice made its way into the lives of most of the world's people. Without the satellite, the

Figure 6.10: Concordance lines for *most of the* in LOCNESS

6.5.1.2.2 Bundles specifying attributes of preceding nouns/entities

Contrary to bundles that specify attributes of following nouns/entities, *people around the world* incorporates an entity (*people*) that is then specified (*around the world*). As mentioned earlier, the bundle *people around the world* (123 times) is the next most frequent bundle after *is one of the*, but the former is distributed more widely (in 107 texts). Figure 6.11 shows 25 of the instances in MCSAW. It is not limited to specifying the number of people that are able to connect with each other internationally, but the bundle also refers to the accessibility and benefits of Facebook, and Facebook as a source of entertainment to many. In addition, it is found that 42 instances of this bundle are used incorrectly by including the numerical classifier *every* as in *every people around the world*. Lines 3-12 also indicate that the bundle is preceded by the determiner *all*, resulting in the 2-word combination *all people*, which can be translated into *semua orang* in the L1. Arguably, it is highly common to use *semua orang* to denote the lexis 'everyone' in the Malay language, and therefore signals a possible L1 transfer. In English, the phrase *all people around the world* and *all the people around the world* are both highly infrequent and hence a-typical: in the 520-million-word Corpus of Contemporary American English (COCA), there are only 3 occurrences of each.



N	Concordance
1	, Yahoo Messenger and particularly Facebook. Based on the statistic, 80% people around the world have Facebook account. I think Facebook has more
2	Facebook in our daily life. This give benefit to us as we can know more about people around the world with Facebook. We can know how they live, what they
3	more advantages than disadvantages because facebook can connect us with all people around the world and we also can learn many things from the facebook.
4	of social networking website known as Facebook are knowledgeable among all people around the world. Especially, the teenagers. Facebook have many
5	you know about Facebook? Facebook is social networks that connects the all people around the world. Facebook is also worldwide, convenient , fast and
6	fast and easy, so facebook is the one of technology that can communicate all people around the world. Facebook has become very important part of our life
7	Advantages and disadvantages of facebook Nowaday, all people around the world like to use technology that can make people work fast
8	is it's the most powerful social media and social networking site for all people around the world. Since Fecebook is networking site, it's not barrier to
9	In conclusion, Facebook can give are negative impact rather than positive to all people around the world that who always acsess this Facebook and misused
10	comfortably and facebook also brings some issues that we can share with all people around the world.Then,we also can learn the new things that will be
11	the creator attract the users. Firstly, Facebook is the easiest way to contact all people around the world with the cheapest cost. We can use chat box to
12	learn many things from the facebook. Firstly,facebook can connect us with all people around the world without limitation because with facebook all people can
13	, twitter and tagged. In the modenisation world right now, I believe that almost people around the world have their own account faceboo. It can give us much
14	value in our society today as we can build a connection between country and people around the world as we are united for a good purposed in life. Lastly,
15	Nowdays, more than 1 billions people around the world have an account for a social network, Facebook. For
16	completely free this makes communication between two or more people cheap. People around the world employ the use of facebook which makes it a
17	in a short period of time. As you can see, Facebook not only can connect people around the world, but they can be a powerful agent to spread news.
18	AGREE OR NOT? Facebook is the most popular social networks that connect people around the world. Everyone have their own Facebook especially students
19	, much time would be wasted. In conclusion, Facebook essentially is to connect people around the world. However, the ethics that should be practiced when
20	Media is having a great revolution in the world to connect people around the world nowadays.As we can see,people in this era more
21	made by the internet.The internet is a wide connection which can connect people around the world.The most famous social networking website right now
22	today is Facebook. Facebook are known as a website or a system that connect people around the world virtually, it has led to many advantages and so as
23	network is currently facebook. Facebook is a social network that connected people around the world. Facebook is also worldwide, coevenient, fast and quick
24	each other, so it can make relationship more closely. In conclusion, connects people around the world is one of the advantage of Facebook. Platform to the
25	there are many social networks that can be found on the Internet that connects people around the world. One of the most popular social networks is Facebook.

Figure 6.11: Concordance lines for *people around the world* in MCSAW

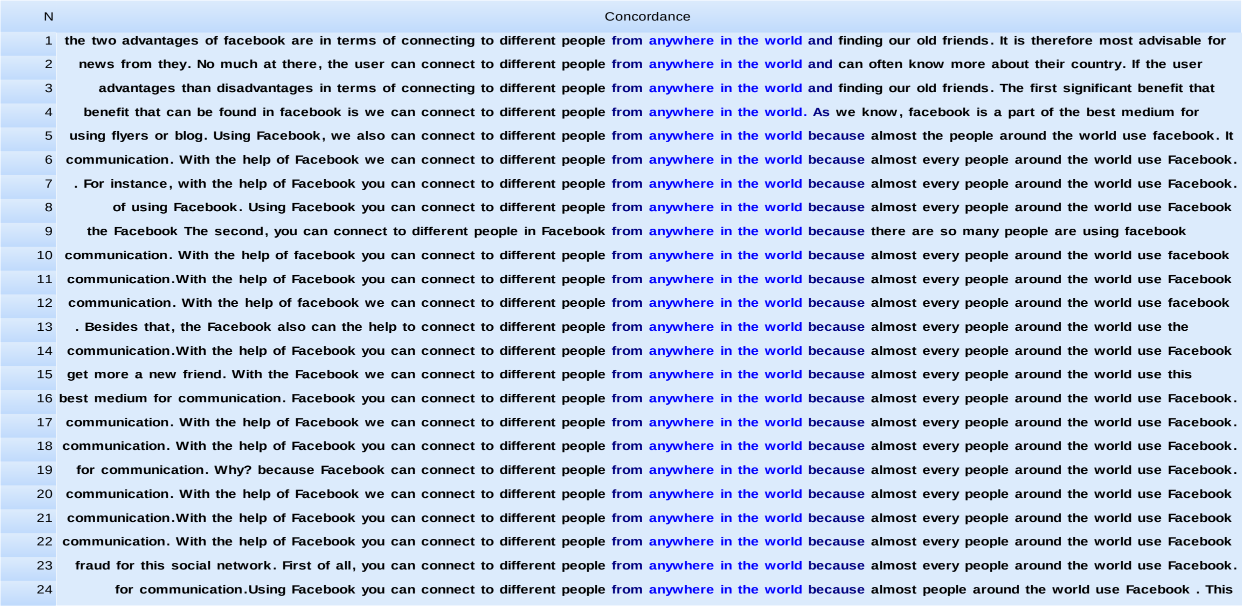
Contrary to the use of *all* in learner texts discussed above, the two instances of *people around the world* in LOCNESS show two distinctive uses in the reference language variety. The first example below shows the bundle occurring after *millions of*. In MCSAW, instances that quantify the bundle aside from the adjective *all*, are *80% people around the world* (line 1), *almost people around the world* (line 13), and *1 billions people around the world* (line 15). The second example taken from LOCNESS shows that *people around the world* co-occurs with *all*, but is then followed by a relative clause that specifies which type of people the writer means (i.e. *all the people…who enjoy watching the sport*).

Exxon's slow response time angered not only the native Alaskans, but millions of *people around the world*. (USARG.txt)

The major influence keeping boxing going is all the *people around the world* who enjoy watching the sport, there is an incredibly big demand for it. (alevels4.txt)

6.5.1.3 Time/place-text-deixis

The recurrent bundle that makes reference to time/place is *anywhere in the world*. It can be seen that, in all 24 instances, the bundle is used in the sentence *with the help of Facebook, we/you can connect to different people from* ***anywhere in the world*** *because almost every people around the world use Facebook* (19 times). The amount of tautology revealed in the concordance lines is likely due to the effects of classroom prompts while writing. These instances, therefore, do not offer us much insight into Malaysian learner language more generally.



N	Concordance
1	the two advantages of facebook are in terms of connecting to different people from anywhere in the world and finding our old friends. It is therefore most advisable for
2	news from they. No much at there, the user can connect to different people from anywhere in the world and can often know more about their country. If the user
3	advantages than disadvantages in terms of connecting to different people from anywhere in the world and finding our old friends. The first significant benefit that
4	benefit that can be found in facebook is we can connect to different people from anywhere in the world. As we know, facebook is a part of the best medium for
5	using flyers or blog. Using Facebook, we also can connect to different people from anywhere in the world because almost the people around the world use facebook. It
6	communication. With the help of Facebook we can connect to different people from anywhere in the world because almost every people around the world use Facebook.
7	. For instance, with the help of Facebook you can connect to different people from anywhere in the world because almost every people around the world use Facebook.
8	of using Facebook. Using Facebook you can connect to different people from anywhere in the world because almost every people around the world use Facebook
9	the Facebook The second, you can connect to different people in Facebook from anywhere in the world because there are so many people are using facebook
10	communication. With the help of facebook you can connect to different people from anywhere in the world because almost every people around the world use facebook
11	communication.With the help of Facebook you can connect to different people from anywhere in the world because almost every people around the world use Facebook
12	communication. With the help of facebook we can connect to different people from anywhere in the world because almost every people around the world use facebook
13	. Besides that, the Facebook also can the help to connect to different people from anywhere in the world because almost every people around the world use the
14	communication.With the help of Facebook you can connect to different people from anywhere in the world because almost every people around the world use Facebook
15	get more a new friend. With the Facebook we can connect to different people from anywhere in the world because almost every people around the world use this
16	best medium for communication. Facebook you can connect to different people from anywhere in the world because almost every people around the world use Facebook.
17	communication. With the help of Facebook we can connect to different people from anywhere in the world because almost every people around the world use Facebook.
18	communication. With the help of Facebook you can connect to different people from anywhere in the world because almost every people around the world use Facebook.
19	for communication. Why? because Facebook can connect to different people from anywhere in the world because almost every people around the world use Facebook.
20	communication. With the help of Facebook we can connect to different people from anywhere in the world because almost every people around the world use Facebook
21	communication.With the help of Facebook you can connect to different people from anywhere in the world because almost every people around the world use Facebook
22	communication. With the help of Facebook you can connect to different people from anywhere in the world because almost every people around the world use Facebook
23	fraud for this social network. First of all, you can connect to different people from anywhere in the world because almost every people around the world use Facebook.
24	for communication.Using Facebook you can connect to different people from anywhere in the world because almost people around the world use Facebook . This

Figure 6.12: Concordance lines for *anywhere in the world* in MCSAW

Contrary to the use of *anywhere in the world* in MCSAW (Figure 6.12), only three instances are found in LOCNESS, which does not make use of this bundle in reference to people, but to other forms of entities. This could be explained as being related to topic variability in the corpus compared to MCSAW.

Britain now has the most cars per mile of road *anywhere in the world* and modern traffic policies are not tackling these problems. (alevels1.txt)

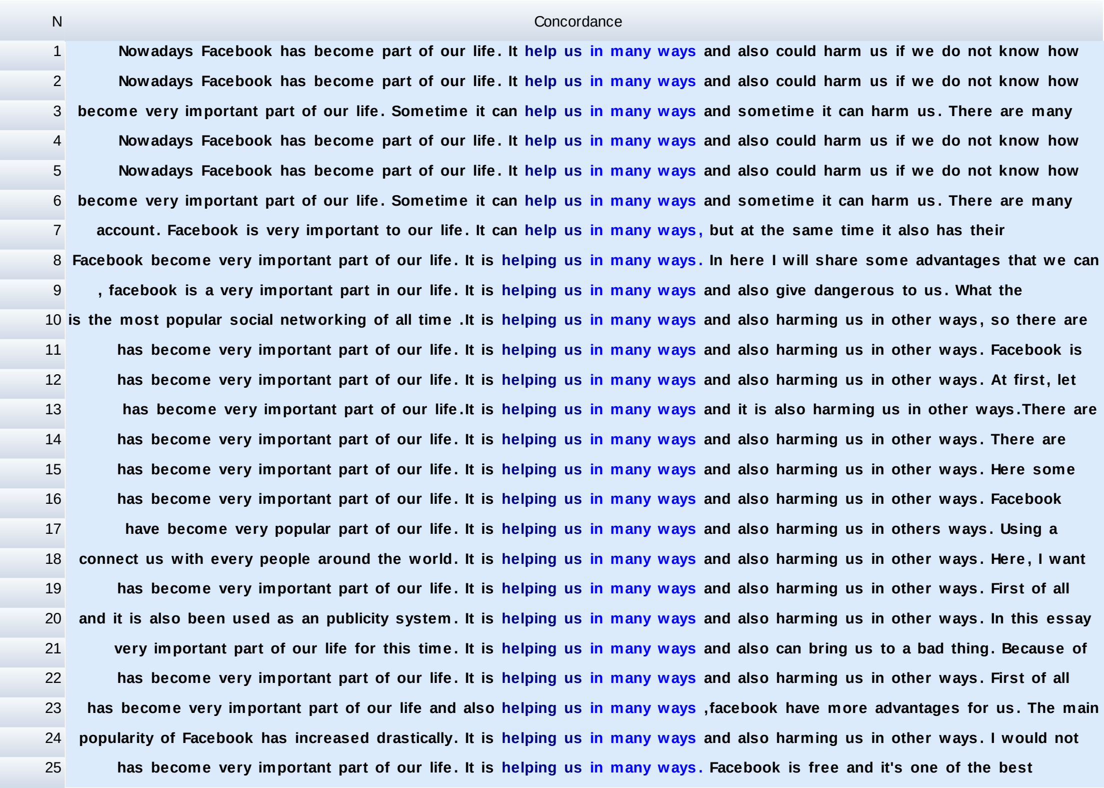
It has given us the freedom to travel anywhere within our own peninsular and, including our travel, almost *anywhere in the world*. (alevels1.txt)

I think the royal family and Monarchy is a a tradition in UK and *anywhere in the world*, UK is known by its royal family, so I do not think it should be abolished. (alevels8.txt)

6.5.1.4 Imprecision bundles

In many ways occurs 65 times in 64 texts in MCSAW. Apart from four instances, it is found that the bundle co-occurs with the word form *help* in 61 instances, as shown in Figure 6.13. 50 of these instances are found to be similar, *It is helping us* ***in many ways*** *and also harming us* *in other ways*. The remaining 11 instances incorporate *it can help*/*it help*(s) *us in many ways*. Thus, majority of the time, this bundle is used to refer to the variety of ways in which Facebook can be seen as helpful. In addition, it is interesting to find that, in many examples, *in many ways* also co-occurs with *in other ways*.

Aside from this, it can be seen that most instances of this bundle are also used in more than one line. Further inspection, however, reveals that the essays are not identical, although some preceding sentences are the same (e.g. lines 11-17): only sections of these essays are identical, rather than the whole essay. As previously argued, it is worth reflecting whether the teacher has 'brainstormed' a few sentences/ideas with learners before making them write the essay, or whether learners were given prompt phrases as writing aids. Another question would be whether this represents large-scale cheating/copying. Given the vast amount of tautological evidence discovered through the analyses, it might also be worth pointing out that these multiple occurrences of repetition pose serious questions about the usefulness of the MCSAW corpus for the analysis of learner language and/or the methodology of using the 'range' function for down-sampling, when a corpus with only limited topic variability is used.



N	Concordance
1	Nowadays Facebook has become part of our life. It help us in many ways and also could harm us if we do not know how
2	Nowadays Facebook has become part of our life. It help us in many ways and also could harm us if we do not know how
3	become very important part of our life. Sometime it can help us in many ways and sometime it can harm us. There are many
4	Nowadays Facebook has become part of our life. It help us in many ways and also could harm us if we do not know how
5	Nowadays Facebook has become part of our life. It help us in many ways and also could harm us if we do not know how
6	become very important part of our life. Sometime it can help us in many ways and sometime it can harm us. There are many
7	account. Facebook is very important to our life. It can help us in many ways, but at the same time it also has their
8	Facebook become very important part of our life. It is helping us in many ways. In here I will share some advantages that we can
9	, facebook is a very important part in our life. It is helping us in many ways and also give dangerous to us. What the
10	is the most popular social networking of all time .It is helping us in many ways and also harming us in other ways, so there are
11	has become very important part of our life. It is helping us in many ways and also harming us in other ways. Facebook is
12	has become very important part of our life. It is helping us in many ways and also harming us in other ways. At first, let
13	has become very important part of our life.It is helping us in many ways and it is also harming us in other ways.There are
14	has become very important part of our life. It is helping us in many ways and also harming us in other ways. There are
15	has become very important part of our life. It is helping us in many ways and also harming us in other ways. Here some
16	has become very important part of our life. It is helping us in many ways and also harming us in other ways. Facebook
17	have become very popular part of our life. It is helping us in many ways and also harming us in others ways. Using a
18	connect us with every people around the world. It is helping us in many ways and also harming us in other ways. Here, I want
19	has become very important part of our life. It is helping us in many ways and also harming us in other ways. First of all
20	and it is also been used as an publicity system. It is helping us in many ways and also harming us in other ways. In this essay
21	very important part of our life for this time. It is helping us in many ways and also can bring us to a bad thing. Because of
22	has become very important part of our life. It is helping us in many ways and also harming us in other ways. First of all
23	has become very important part of our life and also helping us in many ways ,facebook have more advantages for us. The main
24	popularity of Facebook has increased drastically. It is helping us in many ways and also harming us in other ways. I would not
25	has become very important part of our life. It is helping us in many ways. Facebook is free and it's one of the best

Figure 6.13: Concordance lines for *in many ways* in MCSAW

Four instances that do not include the word form *help* in the repeated constructions, as presented above, are shown below. *In many ways* is seen to make ambiguous the particular events: the process of communication among Facebook users, the phenomenon (Facebook), college benefits, and the ability/possibility of searching for friends. It is found that 'many ways' are specified in what follows, notably by words *like*, *either* and the following sentences. In turn, the bundle can be seen to structure the text topically, providing cohesion, and as a result may function as a 'hyper-Theme' (Martin & Rose, 2003). In contrast to 'macro-Themes' explained in Chapter 4, hyper-Themes signify "an introductory sentence or group of sentences in a paragraph which is established to predict a particular pattern of interaction among [lexical/taxonomic] strings, [cohesive] chains and Theme selection in following sentences" (Martin, 1992: p. 437).

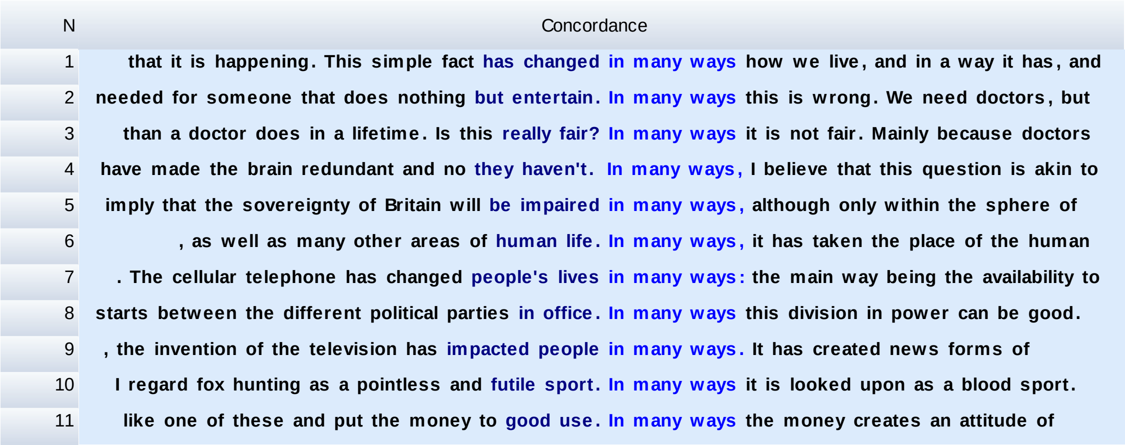
Facebook's user can communicate *in many ways* like sending a massage or making a video call if their friend is online (MCSAW_390.txt)

This phenomenon had brought many changes for everyone *in many ways* either in a good or bad ways (MCSAW_430.txt)

First of all, the college can help these students from not to burden their family *in many ways* (MCSAW_281.txt)

In addition, we can find our old friends *in many ways* (MCSAW_177.txt)

In comparison, 11 instances in LOCNESS, presented in Figure 6.14, also show use of this bundle to indicate various or imprecise details of reasons for a particular argument to be made. Lines 2, 3 and 9, specifically, show that *in many ways* is used to express the writer's disagreement as regards a number of explanations, which are then clarified further (i.e. it also appears to function as a hyper-Theme). It can also be seen that the bundle co-occurs with verbs such as *changed*, *impaired*, and *impacted*, in contrast to learners' overuse of *help*.



N	Concordance
1	that it is happening. This simple fact has changed in many ways how we live, and in a way it has, and
2	needed for someone that does nothing but entertain. In many ways this is wrong. We need doctors, but
3	than a doctor does in a lifetime. Is this really fair? In many ways it is not fair. Mainly because doctors
4	have made the brain redundant and no they haven't. In many ways, I believe that this question is akin to
5	imply that the sovereignty of Britain will be impaired in many ways, although only within the sphere of
6	, as well as many other areas of human life. In many ways, it has taken the place of the human
7	. The cellular telephone has changed people's lives in many ways: the main way being the availability to
8	starts between the different political parties in office. In many ways this division in power can be good.
9	, the invention of the television has impacted people in many ways. It has created news forms of
10	I regard fox hunting as a pointless and futile sport. In many ways it is looked upon as a blood sport.
11	like one of these and put the money to good use. In many ways the money creates an attitude of

Figure 6.14: Concordance lines for *in many ways* in LOCNESS

6.5.1.5 Other referential bundles

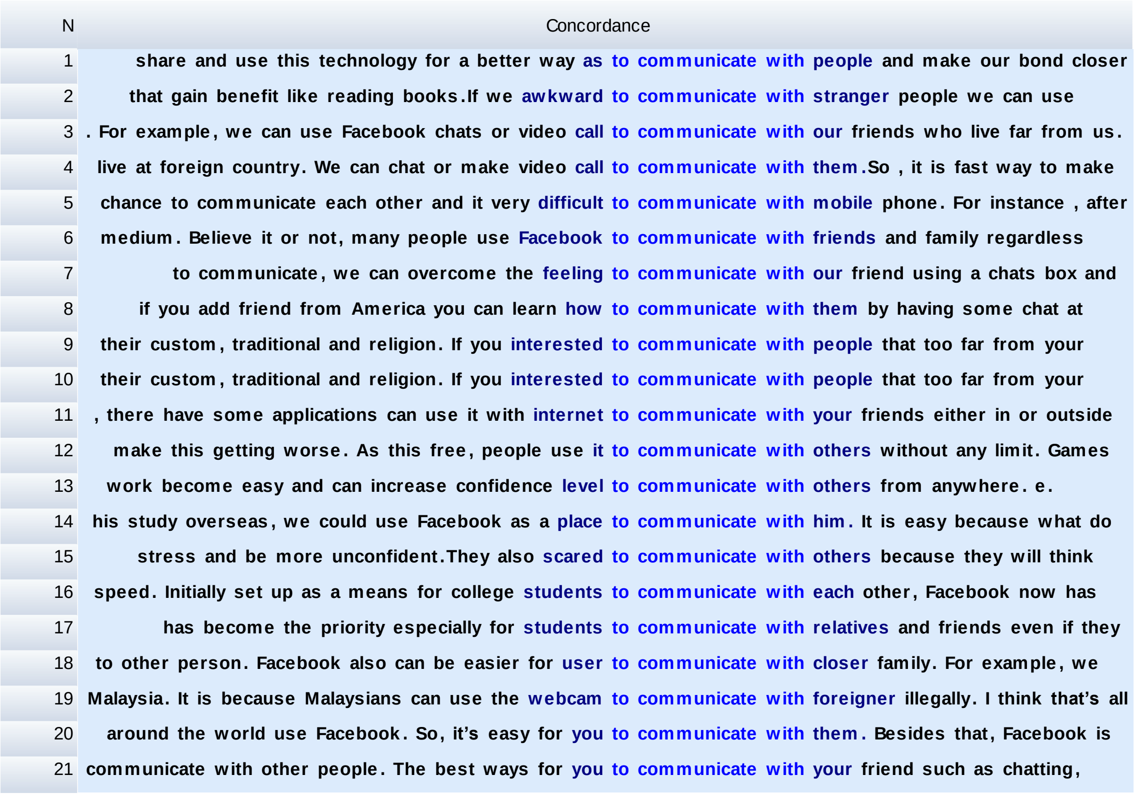
Apart from the above-mentioned referential bundles, three recurrent bundles also make reference to physical or abstract entities, but do not clearly belong to the existing sub-groups. These consist of the bundles *to communicate with* (99), *with their friends* (67), and *for us to* (54).

The bundle *to communicate with* occurs 152 times in MCSAW, and is considered to be the most recurring 3-word bundle in the corpus. It is found that, in 43 occurrences, the bundle co-occurs with the lexis *chance* in *chance to communicate with*. In fact, the same sentence is produced with this bundle, which is *When a friend goes away to any other place, we often don't get the chance* ***to communicate with*** *him or her*. Other similar sentences are found in relation to the bundle such as in *But now Facebook gives us the opportunity* ***to communicate with*** *our old friend very easily without any cost*, and *it gives us the opportunity* ***to communicate with*** *them easily without involving any cost*, which show the bundle to co-occur with *opportunity* 38 times in the corpus. The bundle also co-occurs with *want* (12), *easy* (11), *us* (8), *way* (6), *medium* (4), *people* (4), and *used* (3). Although overuse of frequent sentences may be attributed to the use of prompts in classroom practices (as mentioned earlier), it could also account for why salient bundles like *to communicate with* are significantly over-used. Other instances are shown in Figure 6.15.

In LOCNESS, *to communicate with* is not over-used as it is in MCSAW: only two instances are found, which are shown below. Similarly, the bundle is seen to make reference to physical/abstract entities (i.e. *art* and *device*), which relates to the process of communication expressed by this bundle.

Similarly, art also helps us *to communicate with* each other and show others our values (USARG.txt)

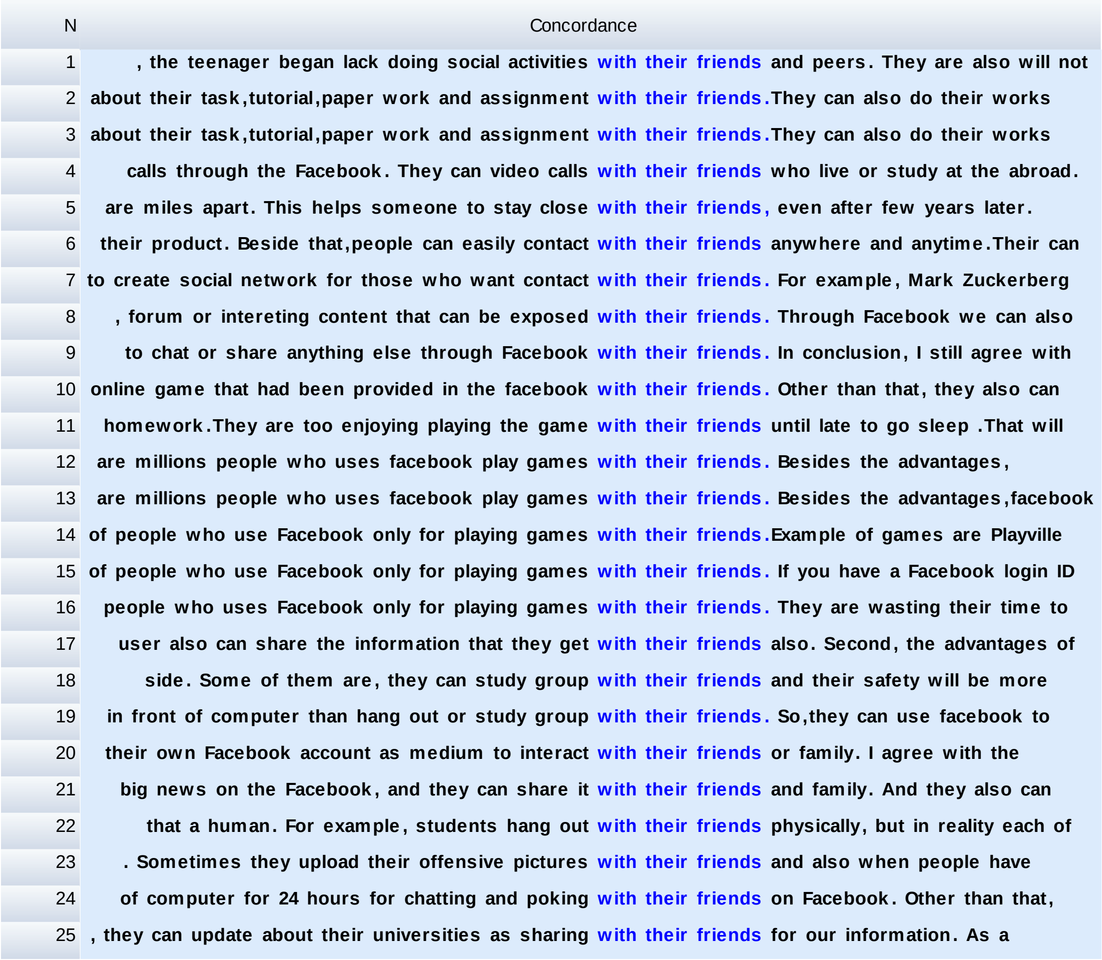
It is also a very influential and powerful device in that it is the easiest way in which *to communicate with* populations (USARG.txt)



N	Concordance
1	share and use this technology for a better way as to communicate with people and make our bond closer
2	that gain benefit like reading books.If we awkward to communicate with stranger people we can use
3	. For example, we can use Facebook chats or video call to communicate with our friends who live far from us.
4	live at foreign country. We can chat or make video call to communicate with them.So , it is fast way to make
5	chance to communicate each other and it very difficult to communicate with mobile phone. For instance , after
6	medium. Believe it or not, many people use Facebook to communicate with friends and family regardless
7	to communicate, we can overcome the feeling to communicate with our friend using a chats box and
8	if you add friend from America you can learn how to communicate with them by having some chat at
9	their custom, traditional and religion. If you interested to communicate with people that too far from your
10	their custom, traditional and religion. If you interested to communicate with people that too far from your
11	, there have some applications can use it with internet to communicate with your friends either in or outside
12	make this getting worse. As this free, people use it to communicate with others without any limit. Games
13	work become easy and can increase confidence level to communicate with others from anywhere. e.
14	his study overseas, we could use Facebook as a place to communicate with him. It is easy because what do
15	stress and be more unconfident.They also scared to communicate with others because they will think
16	speed. Initially set up as a means for college students to communicate with each other, Facebook now has
17	has become the priority especially for students to communicate with relatives and friends even if they
18	to other person. Facebook also can be easier for user to communicate with closer family. For example, we
19	Malaysia. It is because Malaysians can use the webcam to communicate with foreigner illegally. I think that's all
20	around the world use Facebook. So, it's easy for you to communicate with them. Besides that, Facebook is
21	communicate with other people. The best ways for you to communicate with your friend such as chatting,

Figure 6.15: Concordance lines for *to communicate with* in MCSAW

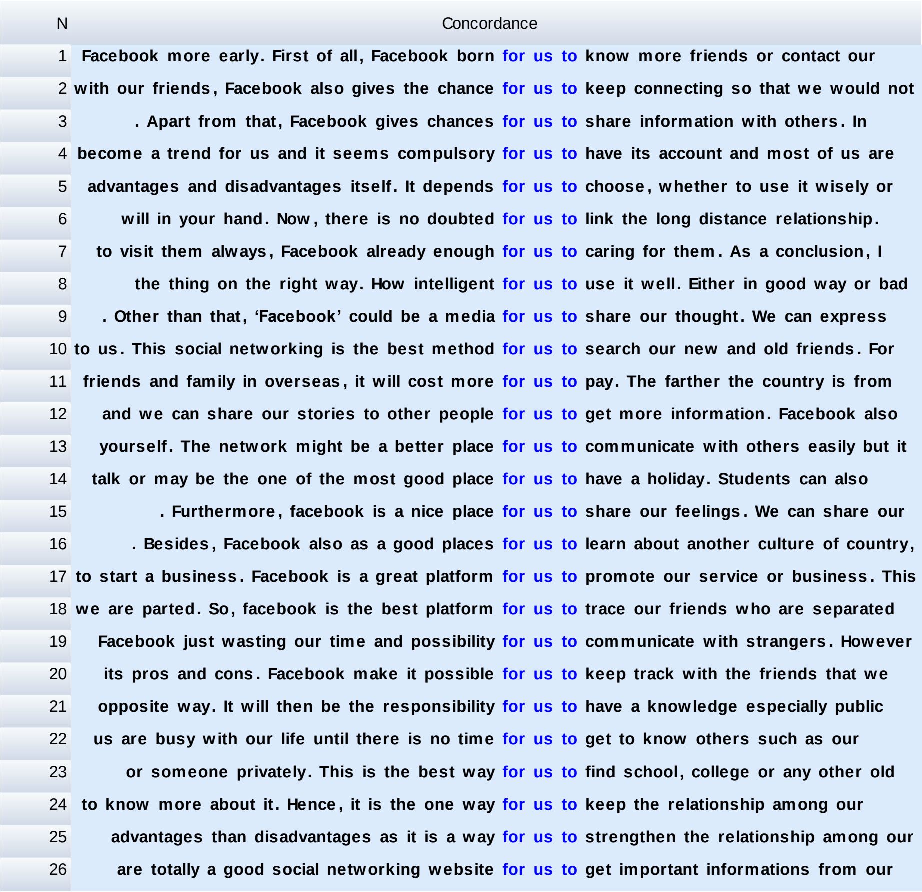
With their friends occurs 89 times and spread across 67 texts in MCSAW. Most instances of this bundle are found to co-occur frequently with *communicate* (19), such as in *They can communicate* ***with their friends*** *without using money and …* (MCSAW_223.txt). The bundle is also found to frequently co-occur with *chat/chatting/chitchatting* (13), such as in *People will chat* ***with their friends*** *and families on Facebook* (MCSAW_176.txt), and with connect*/connected/connecting* (12), in *Once people connect* ***with their friends*** *on facebook* *they will tend to chat …* (MCSAW_239.txt). Other examples are shown in Figure 6.16. Unlike in MCSAW, this bundle occurs only once in LOCNESS, *I know that most people ride in a group* ***with their friends*** *and they bump each other for fun, but it can turn out not to be fun* (USARG.txt).



N	Concordance
1	, the teenager began lack doing social activities with their friends and peers. They are also will not
2	about their task,tutorial,paper work and assignment with their friends.They can also do their works
3	about their task,tutorial,paper work and assignment with their friends.They can also do their works
4	calls through the Facebook. They can video calls with their friends who live or study at the abroad.
5	are miles apart. This helps someone to stay close with their friends, even after few years later.
6	their product. Beside that,people can easily contact with their friends anywhere and anytime.Their can
7	to create social network for those who want contact with their friends. For example, Mark Zuckerberg
8	, forum or intereting content that can be exposed with their friends. Through Facebook we can also
9	to chat or share anything else through Facebook with their friends. In conclusion, I still agree with
10	online game that had been provided in the facebook with their friends. Other than that, they also can
11	homework.They are too enjoying playing the game with their friends until late to go sleep .That will
12	are millions people who uses facebook play games with their friends. Besides the advantages,
13	are millions people who uses facebook play games with their friends. Besides the advantages,facebook
14	of people who use Facebook only for playing games with their friends.Example of games are Playville
15	of people who use Facebook only for playing games with their friends. If you have a Facebook login ID
16	people who uses Facebook only for playing games with their friends. They are wasting their time to
17	user also can share the information that they get with their friends also. Second, the advantages of
18	side. Some of them are, they can study group with their friends and their safety will be more
19	in front of computer than hang out or study group with their friends. So,they can use facebook to
20	their own Facebook account as medium to interact with their friends or family. I agree with the
21	big news on the Facebook, and they can share it with their friends and family. And they also can
22	that a human. For example, students hang out with their friends physically, but in reality each of
23	. Sometimes they upload their offensive pictures with their friends and also when people have
24	of computer for 24 hours for chatting and poking with their friends on Facebook. Other than that,
25	, they can update about their universities as sharing with their friends for our information. As a

Figure 6.16: Concordance lines for *with their friends* in MCSAW

Finally, *for us to* occurs 63 times in MCSAW throughout 54 texts (Figure 6.17). It is found that the bundle frequently co-occurs with *easy/easier* (11), such as in *It is also* ***easy for us to*** *keep in touch with our family members, …* (MCSAW_229.txt), *medium* (9), such as in *One of the advantages of facebook is it can be a* ***medium for us to*** *connect with people all around the world* (MCSAW_81.txt), *opportunity/opportunities* (5), such as in *It also gives an* ***opportunity for us to*** *know more about others custom, culture …* (MCSAW_215.txt), and *good* (4), such as in *Facebook is* ***good for us to*** *stay connected with our friends …* (MCSAW_14.txt). Other frequent collocates are *important* (3) and *impossible* (3). In addition, the bundle *for us to* that incorporates use of the pronoun *us* relates to both the speaker/writer and reader, which is possibly intended to appeal to the strategy of shared experience, as discussed in Chapter 5.



N	Concordance
1	Facebook more early. First of all, Facebook born for us to know more friends or contact our
2	with our friends, Facebook also gives the chance for us to keep connecting so that we would not
3	. Apart from that, Facebook gives chances for us to share information with others. In
4	become a trend for us and it seems compulsory for us to have its account and most of us are
5	advantages and disadvantages itself. It depends for us to choose, whether to use it wisely or
6	will in your hand. Now, there is no doubted for us to link the long distance relationship.
7	to visit them always, Facebook already enough for us to caring for them. As a conclusion, I
8	the thing on the right way. How intelligent for us to use it well. Either in good way or bad
9	. Other than that, 'Facebook' could be a media for us to share our thought. We can express
10	to us. This social networking is the best method for us to search our new and old friends. For
11	friends and family in overseas, it will cost more for us to pay. The farther the country is from
12	and we can share our stories to other people for us to get more information. Facebook also
13	yourself. The network might be a better place for us to communicate with others easily but it
14	talk or may be the one of the most good place for us to have a holiday. Students can also
15	. Furthermore, facebook is a nice place for us to share our feelings. We can share our
16	. Besides, Facebook also as a good places for us to learn about another culture of country,
17	to start a business. Facebook is a great platform for us to promote our service or business. This
18	we are parted. So, facebook is the best platform for us to trace our friends who are separated
19	Facebook just wasting our time and possibility for us to communicate with strangers. However
20	its pros and cons. Facebook make it possible for us to keep track with the friends that we
21	opposite way. It will then be the responsibility for us to have a knowledge especially public
22	us are busy with our life until there is no time for us to get to know others such as our
23	or someone privately. This is the best way for us to find school, college or any other old
24	to know more about it. Hence, it is the one way for us to keep the relationship among our
25	advantages than disadvantages as it is a way for us to strengthen the relationship among our
26	are totally a good social networking website for us to get important informations from our

Figure 6.17: Concordance lines for *for us to* in MCSAW

In LOCNESS, there are 3 instances, and they are also used quite similarly, in that the co-occurring words include *easy*, *important* and *use*. Furthermore, use of this bundle in the reference language variety is also in reference to both speaker and hearer, and hence may suggest an appeal to the reader's shared experience, or creating solidarity with readers, as shown below:

Thus, if our trust is shaken it will not be as easy *for us to* allow physicians to do their jobs (USARG.txt).

It is important *for us to* be able to respect Madonna for her natural beauty just the same as we respect Charles Barkley for his basketball abilities (USARG.txt).

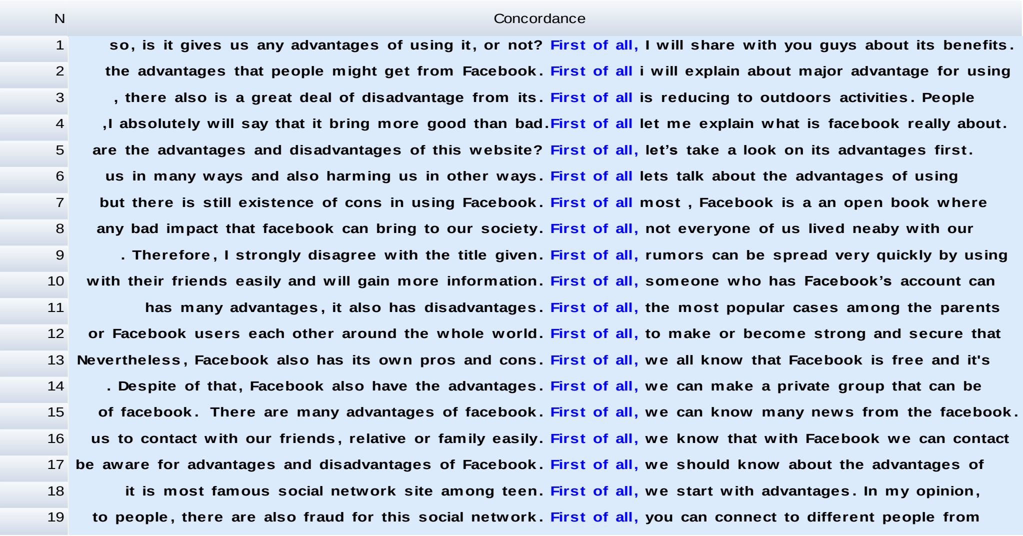
It is *for us to* use, not for it to use us (USARG.txt).

6.5.2 Discourse-organising bundles

As previously mentioned, only shared 3-word bundles have discourse-organising bundles. These include discourse organisers that either introduce a topic (*first of all*, *first and foremost*), elaborate a topic (*is a social*, *have their own*, *for example if*, *then it is*), or make inferences (*as a conclusion*, *it is because*, *this is because*), as is shown in Table A6.2 in the Appendix. The following sub-sections will focus on the recurrent bundles according to their specific sub-categories.

6.5.2.1 Topic introduction/focus

The bundle *first of all* is found 58 times (also throughout 58 texts) in MCSAW and 13 times in LOCNESS. It is described in previous research as generally denoting the rhetorical function of listing items, since this expression is normally used to emphasise the first item of a list (Gilquin & Paquot, 2007: p. 3). This usage is exemplified similarly in both corpora.

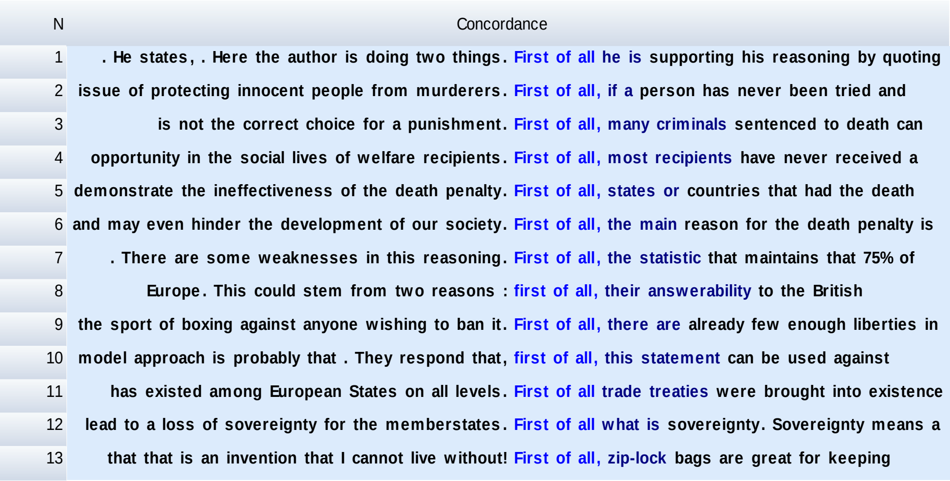


N	Concordance
1	so, is it gives us any advantages of using it, or not? First of all, I will share with you guys about its benefits.
2	the advantages that people might get from Facebook. First of all i will explain about major advantage for using
3	, there also is a great deal of disadvantage from its. First of all is reducing to outdoors activities. People
4	,I absolutely will say that it bring more good than bad.First of all let me explain what is facebook really about.
5	are the advantages and disadvantages of this website? First of all, let’s take a look on its advantages first.
6	us in many ways and also harming us in other ways. First of all lets talk about the advantages of using
7	but there is still existence of cons in using Facebook. First of all most , Facebook is a an open book where
8	any bad impact that facebook can bring to our society. First of all, not everyone of us lived neaby with our
9	. Therefore, I strongly disagree with the title given. First of all, rumors can be spread very quickly by using
10	with their friends easily and will gain more information. First of all, someone who has Facebook’s account can
11	has many advantages, it also has disadvantages. First of all, the most popular cases among the parents
12	or Facebook users each other around the whole world. First of all, to make or become strong and secure that
13	Nevertheless, Facebook also has its own pros and cons. First of all, we all know that Facebook is free and it's
14	. Despite of that, Facebook also have the advantages. First of all, we can make a private group that can be
15	of facebook. There are many advantages of facebook. First of all, we can know many news from the facebook.
16	us to contact with our friends, relative or family easily. First of all, we know that with Facebook we can contact
17	be aware for advantages and disadvantages of Facebook. First of all, we should know about the advantages of
18	it is most famous social network site among teen. First of all, we start with advantages. In my opinion,
19	to people, there are also fraud for this social network. First of all, you can connect to different people from

Figure 6.18: *First of all* concordance lines in MCSAW

In MCSAW (Figure 6.18), *first of all* is found to co-occur with *Facebook* (15), such as in ***First of all***, ***Facebook*** *is free and the best medium for communication regardless age, gender, and religion* (MCSAW_118txt). The bundle is also found co-occurring with the pronoun *it* (3), as anaphoric reference, such as in *It was clearly seen that Facebook bring more harm than good.* ***First of all***, ***it*** *can contribute to health disorders* (MCSAW_380.txt). In addition, some instances of the bundle co-occur with the prompt *advantage*(s), such as in ***First of all***, *the* ***advantages*** *of having Facebook account is we can easily get connect and* …(MCSAW_433.txt). 13 instances, on the other hand, include collocates of the bundle, such as *students*, *college*, *expenses* and *low income parent*, to refer to the hostel topic, such as in ***First of all***, *students can save their expenses* (MCSAW_271.txt) and ***First of all***, *the college can help these students from not to burden their family in many ways* (MCSAW_281.txt). The remaining 19 instances of *first of all*, presented in Figure 6.18, show that the bundle co-occurs with personal pronouns *I*, *you*, *we* and the contraction *let's*.

In contrast, examples of *first of all* in LOCNESS, as shown in Figure 6.19, do not incorporate these features. While there are a considerable number of bundles in LOCNESS, *first of all* is much more frequent (almost 4 times more) in MCSAW, and thus is over-used by learners. One possible explanation lies in the effect of process-based or expository essays taught in schools, with over-teaching of listing signals such as *first of all*, *secondly* and *in conclusion*, at the expense of overt repetition of these items in learner writing. Hyland (1990: p. 72) states that "[t]he shift to a new sequence may be implicit in a topic change, being embedded in the claim, but writers often wish to explicitly guide the reader through the argument stage". He also makes note that students particularly favour using listing signals. In addition, Gilquin and Paquot (2007) argue that *first of all* is more typical of speech than of academic writing, and their overuse in written argumentative essays by Malaysian learners may thus be characterised as somewhat problematic.



N	Concordance
1	. He states, . Here the author is doing two things. First of all he is supporting his reasoning by quoting
2	issue of protecting innocent people from murderers. First of all, if a person has never been tried and
3	is not the correct choice for a punishment. First of all, many criminals sentenced to death can
4	opportunity in the social lives of welfare recipients. First of all, most recipients have never received a
5	demonstrate the ineffectiveness of the death penalty. First of all, states or countries that had the death
6	and may even hinder the development of our society. First of all, the main reason for the death penalty is
7	. There are some weaknesses in this reasoning. First of all, the statistic that maintains that 75% of
8	Europe. This could stem from two reasons : first of all, their answerability to the British
9	the sport of boxing against anyone wishing to ban it. First of all, there are already few enough liberties in
10	model approach is probably that . They respond that, first of all, this statement can be used against
11	has existed among European States on all levels. First of all trade treaties were brought into existence
12	lead to a loss of sovereignty for the memberstates. First of all what is sovereignty. Sovereignty means a
13	that that is an invention that I cannot live without! First of all, zip-lock bags are great for keeping

Figure 6.19: *First of all* concordance lines in LOCNESS

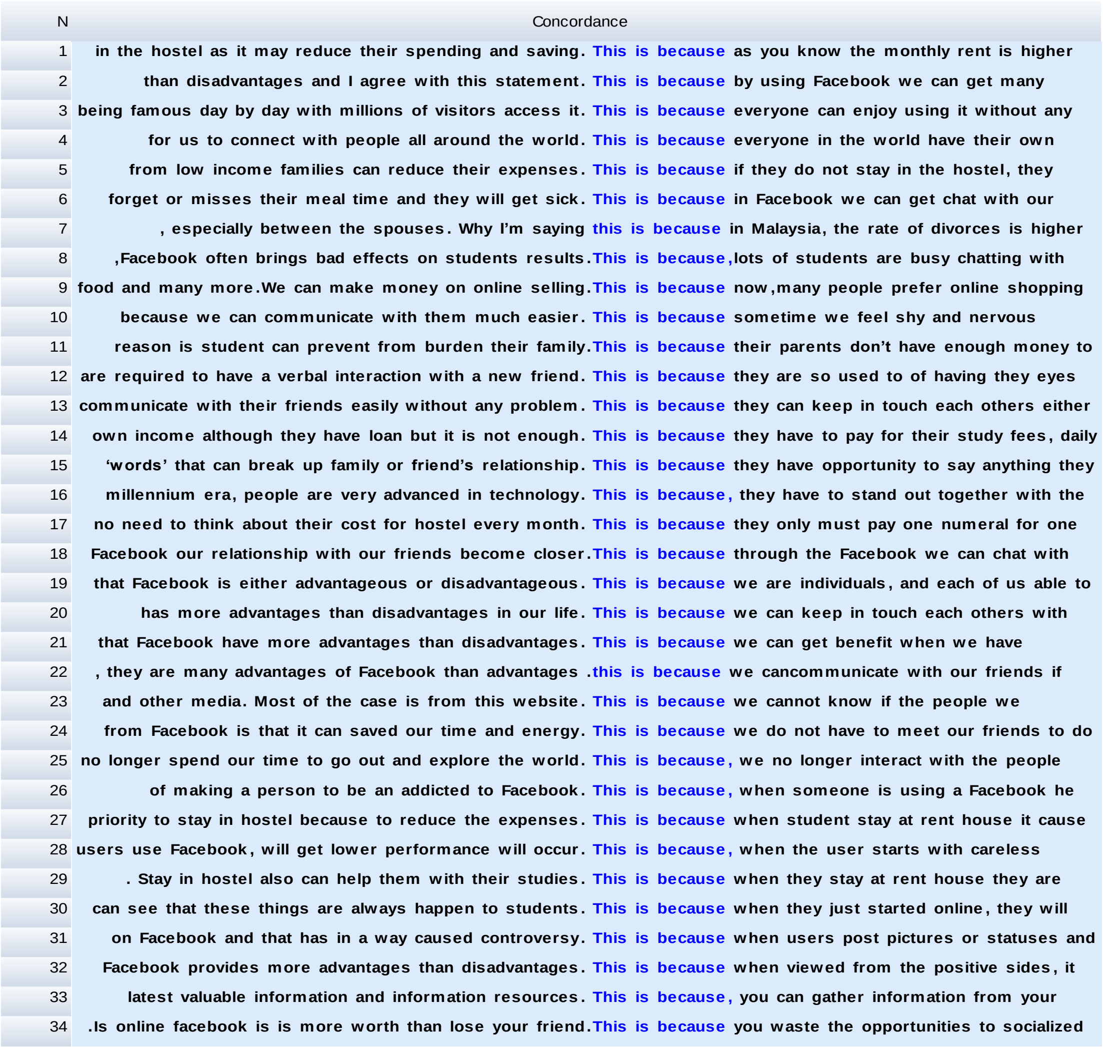
6.5.2.2 Inferential

The bundle *this is because* occurs more often than *first of all*, occurring 98 times, and is distributed throughout 73 texts in MCSAW, but only 12 in LOCNESS. This bundle functions to infer, in that learners suppose a particular claim to be attributed to a certain reason or belief. It also relates to assumption-making that requires the writer to deduce or reason about a particular statement that has been made. This is exemplified in the instances taken from both MCSAW and LOCNESS concordance lines shown below. It can be argued that the two groups of novice writers use this bundle in the same way, but with more occurrences in the learner corpus than in the reference language variety.

However, their rhetorical function (i.e. showing cause and effect) is identified by Gilquin and Paquot (2007)[67] as also being spoken-like, which is similar to the previous bundle, *first of all*. They further add that one way to explain the spoken-like nature of learner writing is by investigating the influence of speech. In the case of Malaysian learners, this is possibly true since *this is because* can be translated into Malay as *ini disebabkan/ini (oleh) kerana*. Interestingly, *ini disebabkan/ini (oleh) kerana* is also found to be used in the same way as *this is because* (i.e. for cause and effect purposes). The 34 instances presented in Figure 6.20 demonstrate uses of this bundle as indicative of oral speech in MCSAW, given the co-

[67] See Table 2.1 for more spoken-like overused lexical items and their rhetorical functions, as taken from Gilquin & Paquot (2007).

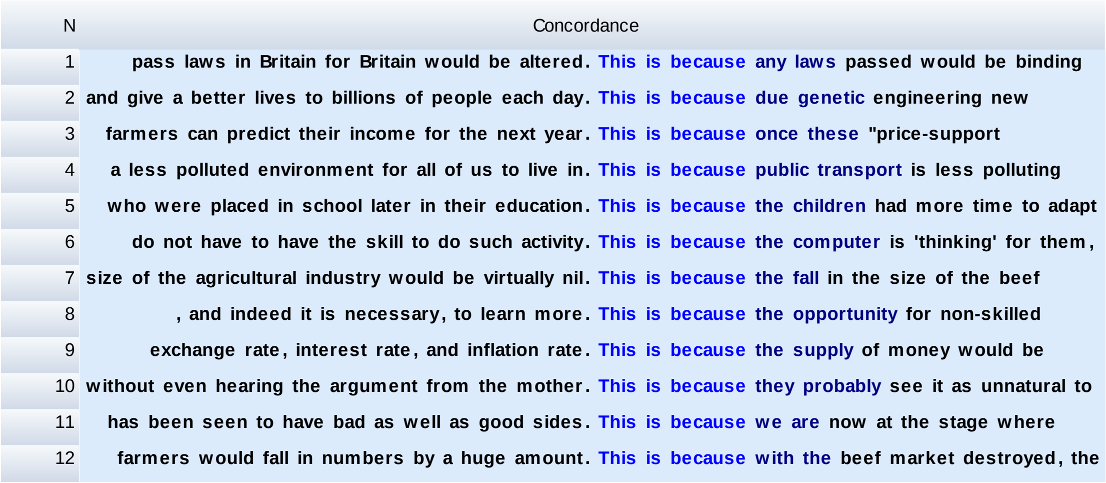
occurrences of personal pronouns as well as speech fragments such as *as you know*, *I'm saying*, and the use of *when*.



N	Concordance
1	in the hostel as it may reduce their spending and saving. This is because as you know the monthly rent is higher
2	than disadvantages and I agree with this statement. This is because by using Facebook we can get many
3	being famous day by day with millions of visitors access it. This is because everyone can enjoy using it without any
4	for us to connect with people all around the world. This is because everyone in the world have their own
5	from low income families can reduce their expenses. This is because if they do not stay in the hostel, they
6	forget or misses their meal time and they will get sick. This is because in Facebook we can get chat with our
7	, especially between the spouses. Why I'm saying this is because in Malaysia, the rate of divorces is higher
8	,Facebook often brings bad effects on students results.This is because,lots of students are busy chatting with
9	food and many more.We can make money on online selling.This is because now,many people prefer online shopping
10	because we can communicate with them much easier. This is because sometime we feel shy and nervous
11	reason is student can prevent from burden their family.This is because their parents don't have enough money to
12	are required to have a verbal interaction with a new friend. This is because they are so used to of having they eyes
13	communicate with their friends easily without any problem. This is because they can keep in touch each others either
14	own income although they have loan but it is not enough. This is because they have to pay for their study fees, daily
15	'words' that can break up family or friend's relationship. This is because they have opportunity to say anything they
16	millennium era, people are very advanced in technology. This is because, they have to stand out together with the
17	no need to think about their cost for hostel every month. This is because they only must pay one numeral for one
18	Facebook our relationship with our friends become closer.This is because through the Facebook we can chat with
19	that Facebook is either advantageous or disadvantageous. This is because we are individuals, and each of us able to
20	has more advantages than disadvantages in our life. This is because we can keep in touch each others with
21	that Facebook have more advantages than disadvantages. This is because we can get benefit when we have
22	, they are many advantages of Facebook than advantages .this is because we cancommunicate with our friends if
23	and other media. Most of the case is from this website. This is because we cannot know if the people we
24	from Facebook is that it can saved our time and energy. This is because we do not have to meet our friends to do
25	no longer spend our time to go out and explore the world. This is because, we no longer interact with the people
26	of making a person to be an addicted to Facebook. This is because, when someone is using a Facebook he
27	priority to stay in hostel because to reduce the expenses. This is because when student stay at rent house it cause
28	users use Facebook, will get lower performance will occur. This is because, when the user starts with careless
29	. Stay in hostel also can help them with their studies. This is because when they stay at rent house they are
30	can see that these things are always happen to students. This is because when they just started online, they will
31	on Facebook and that has in a way caused controversy. This is because when users post pictures or statuses and
32	Facebook provides more advantages than disadvantages. This is because when viewed from the positive sides, it
33	latest valuable information and information resources. This is because, you can gather information from your
34	.Is online facebook is is more worth than lose your friend.This is because you waste the opportunities to socialized

Figure 6.20: *This is because* concordance lines in MCSAW

In contrast, Figure 6.21 shows 12 concordance lines for *this is because* in LOCNESS, in which only one occurrence of the inclusive *we* is found to co-occur immediately after the bundle, in line 11.



N	Concordance
1	pass laws in Britain for Britain would be altered. This is because any laws passed would be binding
2	and give a better lives to billions of people each day. This is because due genetic engineering new
3	farmers can predict their income for the next year. This is because once these "price-support
4	a less polluted environment for all of us to live in. This is because public transport is less polluting
5	who were placed in school later in their education. This is because the children had more time to adapt
6	do not have to have the skill to do such activity. This is because the computer is 'thinking' for them,
7	size of the agricultural industry would be virtually nil. This is because the fall in the size of the beef
8	, and indeed it is necessary, to learn more. This is because the opportunity for non-skilled
9	exchange rate, interest rate, and inflation rate. This is because the supply of money would be
10	without even hearing the argument from the mother. This is because they probably see it as unnatural to
11	has been seen to have bad as well as good sides. This is because we are now at the stage where
12	farmers would fall in numbers by a huge amount. This is because with the beef market destroyed, the

Figure 6.21: *This is because* concordance lines in LOCNESS

6.5.3 Stance bundles

Finally, stance bundles include 4 recurrent shared 4-word bundles and 8 recurrent shared 3-word bundles. Table 6.6 shows that there are few epistemic stance bundles (*to know more about*, *in my opinion*, *as we know*) compared to attitudinal ones. In addition, there are differences between 4-word and 3-word bundles in terms of attitudinal functions. It can also be seen in Table 6.6 that 4-word bundles are mainly used to express three types of attitudinal functions (i.e. ability, importance and emotivity), whereas 3-word bundles are found to be used more variably, namely in expressing desirability, obligation, ability and emotivity. However, there are no bundles showing intention/prediction.

Similar to the previous analyses, recurrent stance bundles (*to know more about*, *in my opinion*, *us the opportunity to*, *that we can*, *important part of our*, *become very important*, *one of the best*, and *the advantages of*) are investigated further in the following sub-sections.

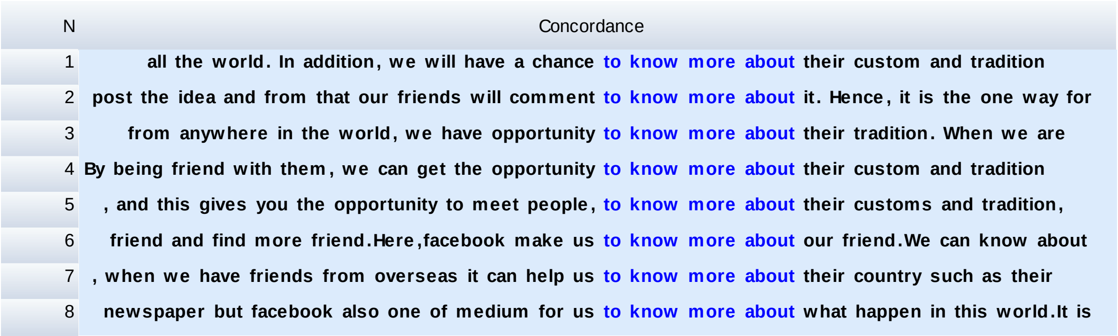
Table 6.6: 4-word and 3-word stance bundles

		4-word bundles	**3-word bundles**
Stance bundles (personal and impersonal, Self/Other	Epistemic	***to know more about*** **(57)**	***in my opinion*** **(82),** *as we know* (46)
	Attitudinal/modality stance		
	Desire	-	*we want to* (38), *have decided to* (31), *you want to* (25)
	Obligation/directive	-	*we have to* (26), *just need to* (30)
	Intention/prediction	-	-
	Ability	*can help us to* (23), ***us the opportunity to*** **(63),** *with the help of* (43)	*you can use* (29), ***that we can*** **(82),** *we can find* (42), *also can be* (34), *can make us* (26), *can be a* (49), *it also can* (26), *the chance to* (48), *help us to* (44), *we can see* (35), *so we can* (32)
	Importance	***important part of our*** **(51)**	***become very important*** **(52)**
	Emotivity	*the best way to* (35), *is the best way* (31), ***one of the best*** **(69),** *this is the best* (22), *it is the best* (31), *of advantages and disadvantages* (21), *there are many advantages* (32)	***the advantages of*** **(74),** *the advantage of* (27), *the disadvantages of* (40), *pros and cons* (36), *they are too* (30), *it is easier* (27), *in the right* (31)

6.5.3.1 Epistemic stance bundles

As a brief reminder, although epistemic stance bundles traditionally express meanings such as certainty or uncertainty, they are conceptualised here to contain reference to the status of information (which is sometimes called *evidentiality*), which includes opinions as well as knowledge about something. This means including bundles with the verb *know* itself. The bundle *to know more about* occurs 57 times in MCSAW, but only once in the reference language variety (in the form of a rhetorical question: *Is it because talk shows show a part of the world they do not understand and are not willing to know more about?* (USARG.txt). In addition, 49 occurrences of this bundle show repetitive use of the verb phrase *gives us the* *opportunity to know more about*, such as in *Thus, it* ***gives us the opportunity to know more about*** *customs and traditions, cultures, religions around the world* (MCSAW_26.txt). As previously highlighted, this phrase adds to the argument that certain bundles are over-used because of essay prompts in classroom teaching or plagiarism on the part of the learners.

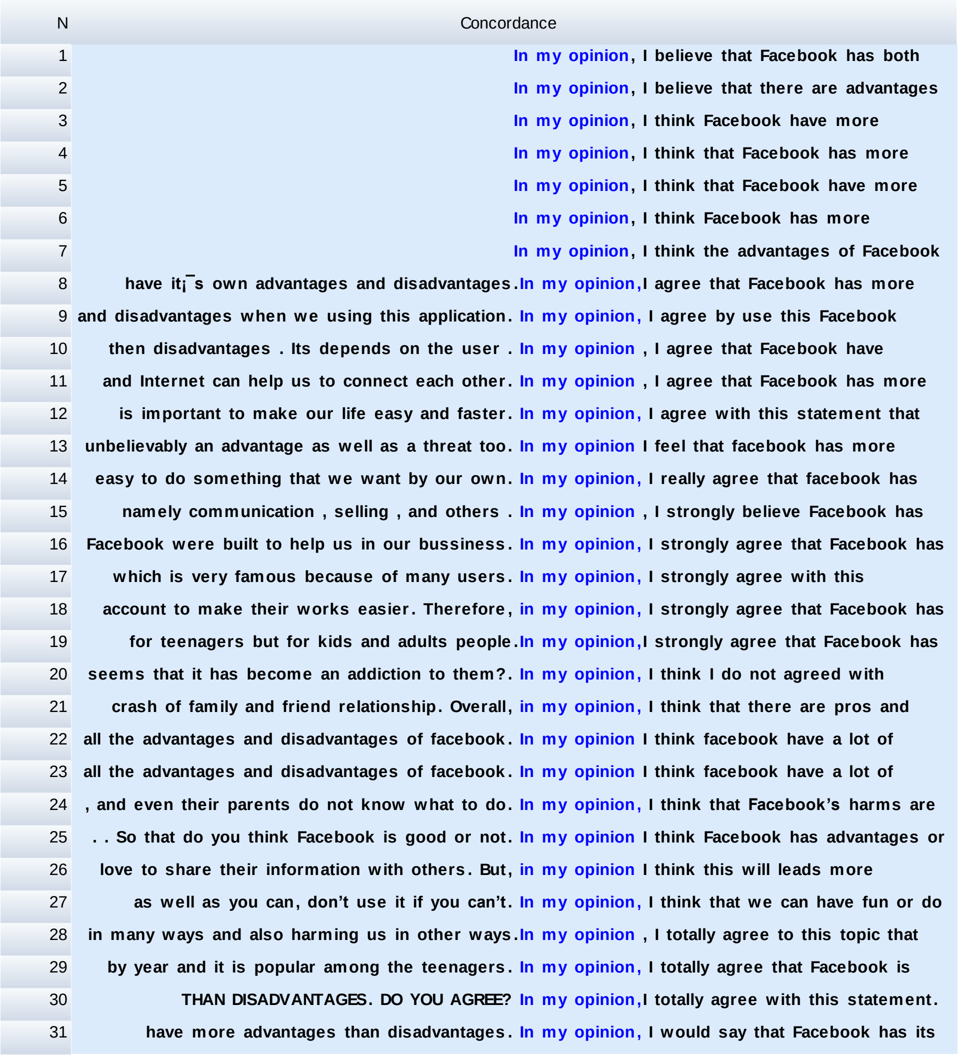
Besides this, most of these examples indicate personal stance, in which first-person plural pronouns *we* and *us* are used to refer to the speaker/writer as well as achieving solidarity with the reader. This can be seen in 8 bundles presented in Figure 6.22, which do not incorporate the repeated lines mentioned above. In most cases, it is a first-person plural pronoun which is the subject of the verb *know*. Only one instance is shown to indicate impersonal stance in MCSAW, *Indirectly, stalkers or scammers can using this method* ***to know more about*** *the person they admire* (MCSAW_255.txt). These examples also show the dependency of the bundle upon preceding co-text (*use X to; a chance to; will comment [in order] to; we have/we can get/gives you (the) opportunity to; make us to; help us to; medium for us to*), including syntactic and collocational errors such as *make us to*.



N	Concordance
1	all the world. In addition, we will have a chance to know more about their custom and tradition
2	post the idea and from that our friends will comment to know more about it. Hence, it is the one way for
3	from anywhere in the world, we have opportunity to know more about their tradition. When we are
4	By being friend with them, we can get the opportunity to know more about their custom and tradition
5	, and this gives you the opportunity to meet people, to know more about their customs and tradition,
6	friend and find more friend.Here,facebook make us to know more about our friend.We can know about
7	, when we have friends from overseas it can help us to know more about their country such as their
8	newspaper but facebook also one of medium for us to know more about what happen in this world.It is

Figure 6.22: Concordance lines for *to know more about* in MCSAW

In addition to the bundle *to know more about*, there are 87 occurrences of *in my opinion* in MCSAW and 28 in LOCNESS. According to Gilquin and Paquot (2007), use of this bundle in this sense makes learners particularly visible as writers. They also note that the use of these expressions is more frequent in speech, and thus contribute to the oral tone of learners' essays (Gilquin & Paquot, 2007). Figure 6.23 shows 31 instances of the bundle in MCSAW as expressing personal epistemic stance and co-occurrences with another personal stance marker (e.g. *I believe, I think*), contributing to double marking that seems tautological, i.e. over-emphasising subjectivity. Furthermore, verbs are also seen to be amplified by emphasisers such as *strongly* (5) and *totally* (3) in lines 15-19 and 28-30. Examples in LOCNESS, as shown in Figure 6.24, on the other hand, do not demonstrate these features. This suggests that learners over-use these combinations more frequently in expressing epistemic stance, and in turn, increase writer-visibility as well as spoken features in learner writing overall.



N	Concordance
1	In my opinion, I believe that Facebook has both
2	In my opinion, I believe that there are advantages
3	In my opinion, I think Facebook have more
4	In my opinion, I think that Facebook has more
5	In my opinion, I think that Facebook have more
6	In my opinion, I think Facebook has more
7	In my opinion, I think the advantages of Facebook
8	have it¡¯s own advantages and disadvantages.In my opinion,I agree that Facebook has more
9	and disadvantages when we using this application. In my opinion, I agree by use this Facebook
10	then disadvantages . Its depends on the user . In my opinion , I agree that Facebook have
11	and Internet can help us to connect each other. In my opinion , I agree that Facebook has more
12	is important to make our life easy and faster. In my opinion, I agree with this statement that
13	unbelievably an advantage as well as a threat too. In my opinion I feel that facebook has more
14	easy to do something that we want by our own. In my opinion, I really agree that facebook has
15	namely communication , selling , and others . In my opinion , I strongly believe Facebook has
16	Facebook were built to help us in our bussiness. In my opinion, I strongly agree that Facebook has
17	which is very famous because of many users. In my opinion, I strongly agree with this
18	account to make their works easier. Therefore, in my opinion, I strongly agree that Facebook has
19	for teenagers but for kids and adults people.In my opinion,I strongly agree that Facebook has
20	seems that it has become an addiction to them?. In my opinion, I think I do not agreed with
21	crash of family and friend relationship. Overall, in my opinion, I think that there are pros and
22	all the advantages and disadvantages of facebook. In my opinion I think facebook have a lot of
23	all the advantages and disadvantages of facebook. In my opinion I think facebook have a lot of
24	, and even their parents do not know what to do. In my opinion, I think that Facebook's harms are
25	. . So that do you think Facebook is good or not. In my opinion I think Facebook has advantages or
26	love to share their information with others. But, in my opinion I think this will leads more
27	as well as you can, don't use it if you can't. In my opinion, I think that we can have fun or do
28	in many ways and also harming us in other ways.In my opinion , I totally agree to this topic that
29	by year and it is popular among the teenagers. In my opinion, I totally agree that Facebook is
30	THAN DISADVANTAGES. DO YOU AGREE? In my opinion,I totally agree with this statement.
31	have more advantages than disadvantages. In my opinion, I would say that Facebook has its

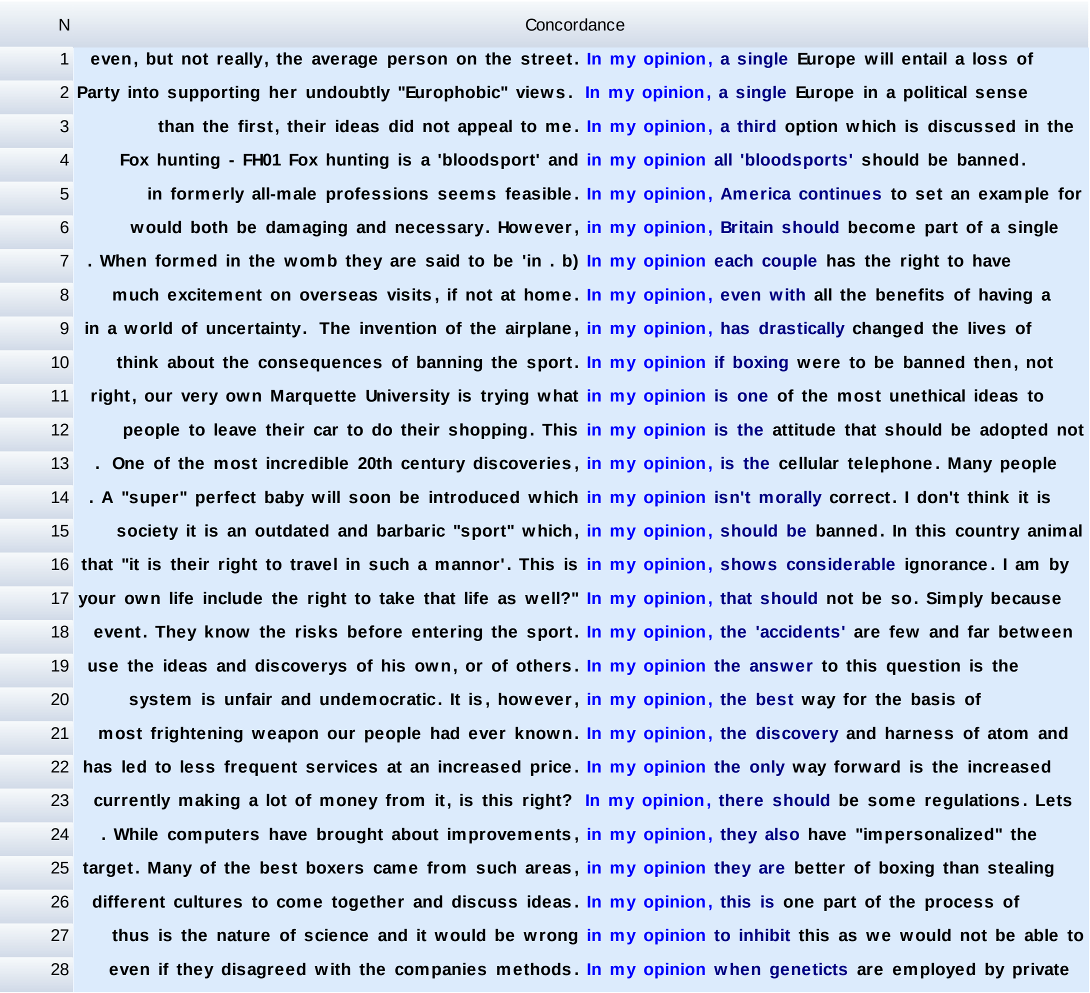
Figure 6.23: *In my opinion* concordance lines in MCSAW

In contrast, closer examination of the concordance lines in Figure 6.24 identifies eight uses of the modal verbs, *should/shouldn't* (6) and *would* (2), in relation to *in my opinion* in LOCNESS. For example, *it would be wrong* ***in my opinion*** *to inhibit this as we would not be able to enjoy the benefits that science can provide…* (line 27) and *This* ***in my opinion*** *is the attitude that should be adopted* (line 12). Only one occurrence of *would* can be found in MCSAW, which is in ***In my opinion, I would say*** *that Facebook has its own equal advantages and disadvantages*; but this, again, demonstrates a double use of personal stance markers, since *I would say that* could be paraphrased as *in my opinion*.

Another difference found in LOCNESS is the co-occurrence of bundle *in my opinion* with *however* to signal opposing views of the writer. For example, *However*, ***in my opinion***, *Britain should become…* (line 6) and *It is, however,* ***in my opinion***, *the best way for the basis of government* (line 20). Learner writers also demonstrate four instances of expressing opposing views in this way, with the incorporation of *however* only once in *However* ***in my opinion***, *it has far greater advantages then the disadvantages* (MCSAW_1txt). The remaining instances are constructed with words *but* and *while*, as exemplified below. These instances are not found in the reference language variety, and in turn, further add to the spoken-like feature of learner writing.

But, in my opinion I think this will leads more disadvantages and less advantages of having Facebook account. (433txt_MCSAW)

Some said that Facebook has weak privacy setting and it will result in the leakage of our private information. *While, in my opinion*, as long as the users know how to use Facebook correctly and accurately, they would not face this problem. (424txt_MCSAW)



N	Concordance
1	even, but not really, the average person on the street. In my opinion, a single Europe will entail a loss of
2	Party into supporting her undoubtly "Europhobic" views. In my opinion, a single Europe in a political sense
3	than the first, their ideas did not appeal to me. In my opinion, a third option which is discussed in the
4	Fox hunting - FH01 Fox hunting is a 'bloodsport' and in my opinion all 'bloodsports' should be banned.
5	in formerly all-male professions seems feasible. In my opinion, America continues to set an example for
6	would both be damaging and necessary. However, in my opinion, Britain should become part of a single
7	. When formed in the womb they are said to be 'in . b) In my opinion each couple has the right to have
8	much excitement on overseas visits, if not at home. In my opinion, even with all the benefits of having a
9	in a world of uncertainty. The invention of the airplane, in my opinion, has drastically changed the lives of
10	think about the consequences of banning the sport. In my opinion if boxing were to be banned then, not
11	right, our very own Marquette University is trying what in my opinion is one of the most unethical ideas to
12	people to leave their car to do their shopping. This in my opinion is the attitude that should be adopted not
13	. One of the most incredible 20th century discoveries, in my opinion, is the cellular telephone. Many people
14	. A "super" perfect baby will soon be introduced which in my opinion isn't morally correct. I don't think it is
15	society it is an outdated and barbaric "sport" which, in my opinion, should be banned. In this country animal
16	that "it is their right to travel in such a mannor'. This is in my opinion, shows considerable ignorance. I am by
17	your own life include the right to take that life as well?" In my opinion, that should not be so. Simply because
18	event. They know the risks before entering the sport. In my opinion, the 'accidents' are few and far between
19	use the ideas and discoverys of his own, or of others. In my opinion the answer to this question is the
20	system is unfair and undemocratic. It is, however, in my opinion, the best way for the basis of
21	most frightening weapon our people had ever known. In my opinion, the discovery and harness of atom and
22	has led to less frequent services at an increased price. In my opinion the only way forward is the increased
23	currently making a lot of money from it, is this right? In my opinion, there should be some regulations. Lets
24	. While computers have brought about improvements, in my opinion, they also have "impersonalized" the
25	target. Many of the best boxers came from such areas, in my opinion they are better of boxing than stealing
26	different cultures to come together and discuss ideas. In my opinion, this is one part of the process of
27	thus is the nature of science and it would be wrong in my opinion to inhibit this as we would not be able to
28	even if they disagreed with the companies methods. In my opinion when genetics are employed by private

Figure 6.24: *In my opinion* concordance lines in LOCNESS

Essentially, the mastering of epistemic devices helps writers to negotiate views/ideas and qualify claims at an appropriate level of commitment (McEnery & Kifle, 2002). The objectives of writing argumentative essays expect learners not only to show their language competence but also their rhetorical skills in writing. As argued by McEnery and Kifle (2002: p. 183),

> [k]nowledge of the types of epistemic modality and the style of their presentation is important for second language writers. These help them to have at their disposal a repertoire of devices that allow them to make claims with the exact degree of certainty or doubt that they intend. It also allows them to achieve native-like competence.

However, as can be seen in the discussion of *in my opinion*, it could be argued that learners will find it more difficult to attain this skill if they are not shown or made aware of undesirable spoken features in writing. In addition, the tautological combination of two personal stance markers, which over-emphasises the subjectivity of the writer's opinion in learner writing, is found to be ubiquitous. Furthermore, examples of prompts in the writing classroom do not appear to be beneficial for learners, since they are easily over-used by learners as a safety net. In its place, pedagogy based on corpora should thus be promoted, so as to provide students with an array of bundle uses, particularly those of epistemic stance, which can minimise repetition in learner writing.

6.5.3.2 Attitudinal/modality stance bundles

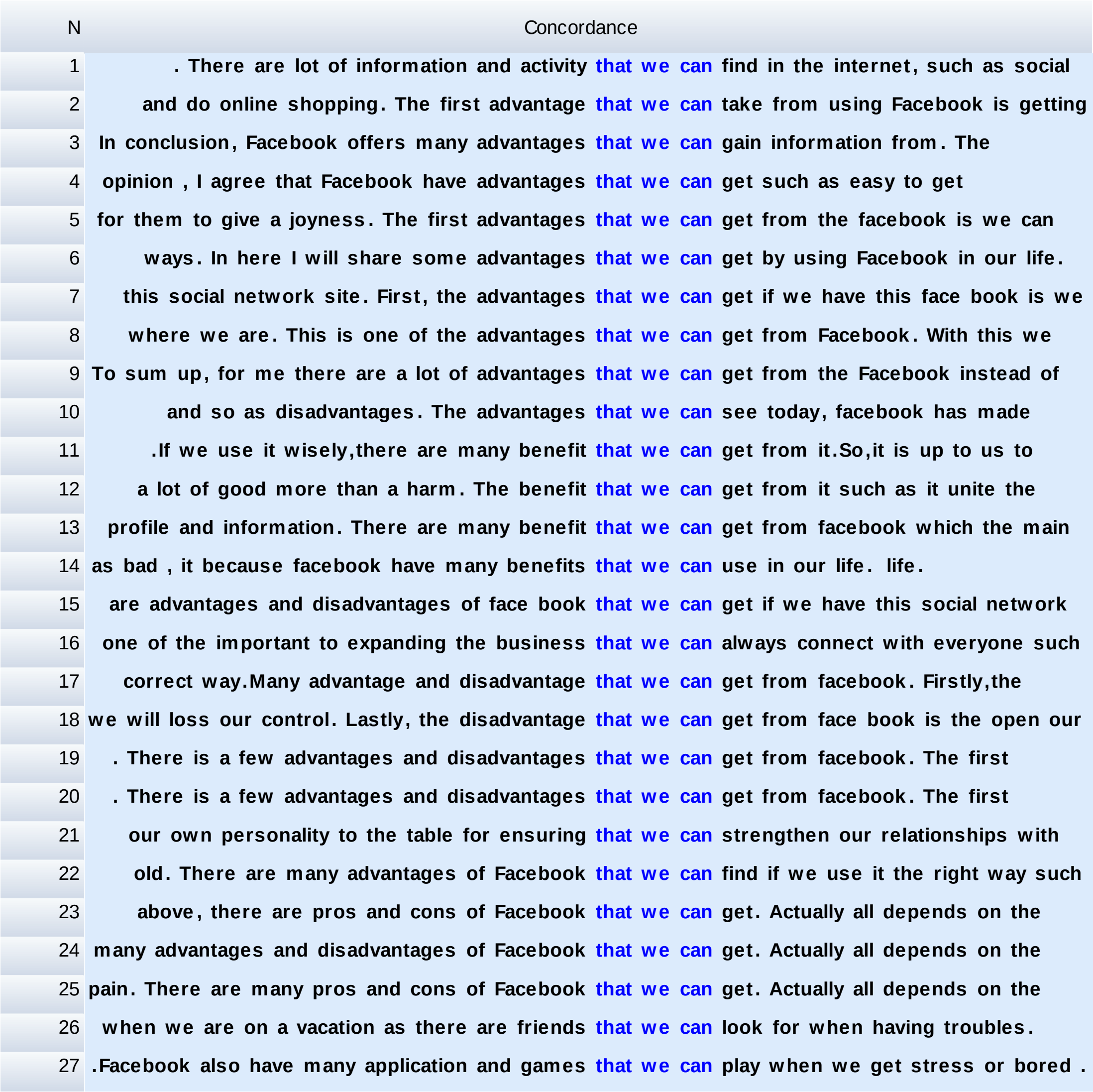
To reiterate, attitudinal or modality stance bundles as conceptualised in this study include: bundles that express a sense of desirability, consisting of bundles that incorporate words such as *want* and *decide* (*we want to, have decided to*, *you want to*); bundles that indicate a sense of obligation or necessity or act as directives (*we have to, just need to*); and bundles expressing ability, with bundles that incorporate the word *can* (e.g. *can help us to, you can use*) and words referring to *opportunity*, *chance* and *help* (e.g. *us the opportunity to*, *with the help of, the chance to*). In addition, bundles that show importance (*important part of our*, *become very important*), and emotivity (e.g. *the best way to, it is easier*) are also included.

6.5.3.2.1 Ability

Two recurrent bundles are identified as expressing ability, namely *us the opportunity to* (63) and *that we can* (82). As previously mentioned, these bundles include the incorporation of the words *can* and *opportunity*, which describe the sense of ability or possibility of something to happen. As regards the bundle *us the opportunity to*, this bundle will not be discussed further because it primarily occurs in the two sentences that have been mentioned in Section 6.5.1.5 (*But now Facebook gives* ***us the opportunity to communicate with*** *our old friend very easily without any cost*, and *it gives* ***us the opportunity to communicate with*** *them easily without involving any cost*), and does not convey much about Malaysian learner language more generally.

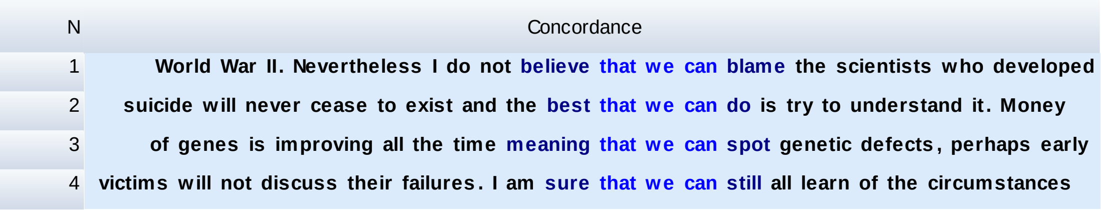
For ease of reading, Figure 6.25 shows 27 of the 51 occurrences of the bundle *that we can* in MCSAW. It can be seen that the bundle mostly refers to the advantages/benefits and/or disadvantages of Facebook in various types of phrases, as has already become apparent in the previous chapter: for example, ADJ + *advantage*(s) + ***that we can*** *take*/*gain*/*get, Facebook have* + *advantages* + ***that we can*** *get, the advantages/disadvantage* + ***that we can*** *get*/*see,* and *there are* (many, a lot of)/*this is one of the* + *advantage*(s)/*benefit*(s) + ***that we can*** *take*/*gain*/*get*/*use*/*see*/*find*. The bundle expresses the reception of these benefits or drawbacks through the ability to *find, take, gain, get, see, use, connect, strengthen, look* and *play*. More specifically, it is found that the '*we* + *can*' construction co-occurs with these verbs in relation to the advantages or disadvantages of Facebook. Apart from the inclusive *we* in these instances, use of high-frequency common words such as *find*, *get* and *see* contributes to learners' writing sounding more spoken-like and thus, are undesirable in academic-style writing (Lee & Chen, 2009). In contrast, Figure 6.26 presents 4 instances of *that we can* in LOCNESS. Lines 1 and 4 explicitly show the writer's expression of personal stance by the use of *I do not believe* and *I am sure*. These instances are not found in MCSAW except for few occurrences such as *I agree*, shown in line 4 in Figure 6.25.

In COCA, however, bundle *that we can* is found to co-occur 1678 times with the adverbial *so* as in *so that we can* + V (e.g. *This article is intended to shine a light on these risks* ***so that we can*** *all be more critical consumers of systematic reviews*). Other frequent (more than 100 times) collocational patterns involving this bundle include words 'hope' (*The* <u>*hope*</u> *is* ***that we can*** *build a social world marked by cooperation and peace*), 'believe' (*While these requests are unlikely to disappear, I* <u>*believe*</u> ***that we can*** *approach these situations with integrity*), 'understand' (*It follows from this* ***that we can*** <u>*understand*</u> *how the Crucifixion is related to the logic of retributive justice*), and 'ways' (*It is essential that future research examine* <u>*ways*</u> ***that we can*** *best support these children and their families*). Although both groups of novice writers do not use this bundle with the consequential meaning (*so that we can*), it can be seen that expressions of personal stance are found (as exemplified above). In spite of this, learners over-use simple common verbs (e.g. *find*, *get*, *see*) with the bundle compared to verbs that mostly discern stance (*hope, believe*, *understand*).



N	Concordance
1	. There are lot of information and activity that we can find in the internet, such as social
2	and do online shopping. The first advantage that we can take from using Facebook is getting
3	In conclusion, Facebook offers many advantages that we can gain information from. The
4	opinion , I agree that Facebook have advantages that we can get such as easy to get
5	for them to give a joyness. The first advantages that we can get from the facebook is we can
6	ways. In here I will share some advantages that we can get by using Facebook in our life.
7	this social network site. First, the advantages that we can get if we have this face book is we
8	where we are. This is one of the advantages that we can get from Facebook. With this we
9	To sum up, for me there are a lot of advantages that we can get from the Facebook instead of
10	and so as disadvantages. The advantages that we can see today, facebook has made
11	.If we use it wisely,there are many benefit that we can get from it.So,it is up to us to
12	a lot of good more than a harm. The benefit that we can get from it such as it unite the
13	profile and information. There are many benefit that we can get from facebook which the main
14	as bad , it because facebook have many benefits that we can use in our life. life.
15	are advantages and disadvantages of face book that we can get if we have this social network
16	one of the important to expanding the business that we can always connect with everyone such
17	correct way.Many advantage and disadvantage that we can get from facebook. Firstly,the
18	we will loss our control. Lastly, the disadvantage that we can get from face book is the open our
19	. There is a few advantages and disadvantages that we can get from facebook. The first
20	. There is a few advantages and disadvantages that we can get from facebook. The first
21	our own personality to the table for ensuring that we can strengthen our relationships with
22	old. There are many advantages of Facebook that we can find if we use it the right way such
23	above, there are pros and cons of Facebook that we can get. Actually all depends on the
24	many advantages and disadvantages of Facebook that we can get. Actually all depends on the
25	pain. There are many pros and cons of Facebook that we can get. Actually all depends on the
26	when we are on a vacation as there are friends that we can look for when having troubles.
27	.Facebook also have many application and games that we can play when we get stress or bored .

Figure 6.25: Concordance lines for *that we can* in MCSAW



N	Concordance
1	World War II. Nevertheless I do not believe that we can blame the scientists who developed
2	suicide will never cease to exist and the best that we can do is try to understand it. Money
3	of genes is improving all the time meaning that we can spot genetic defects, perhaps early
4	victims will not discuss their failures. I am sure that we can still all learn of the circumstances

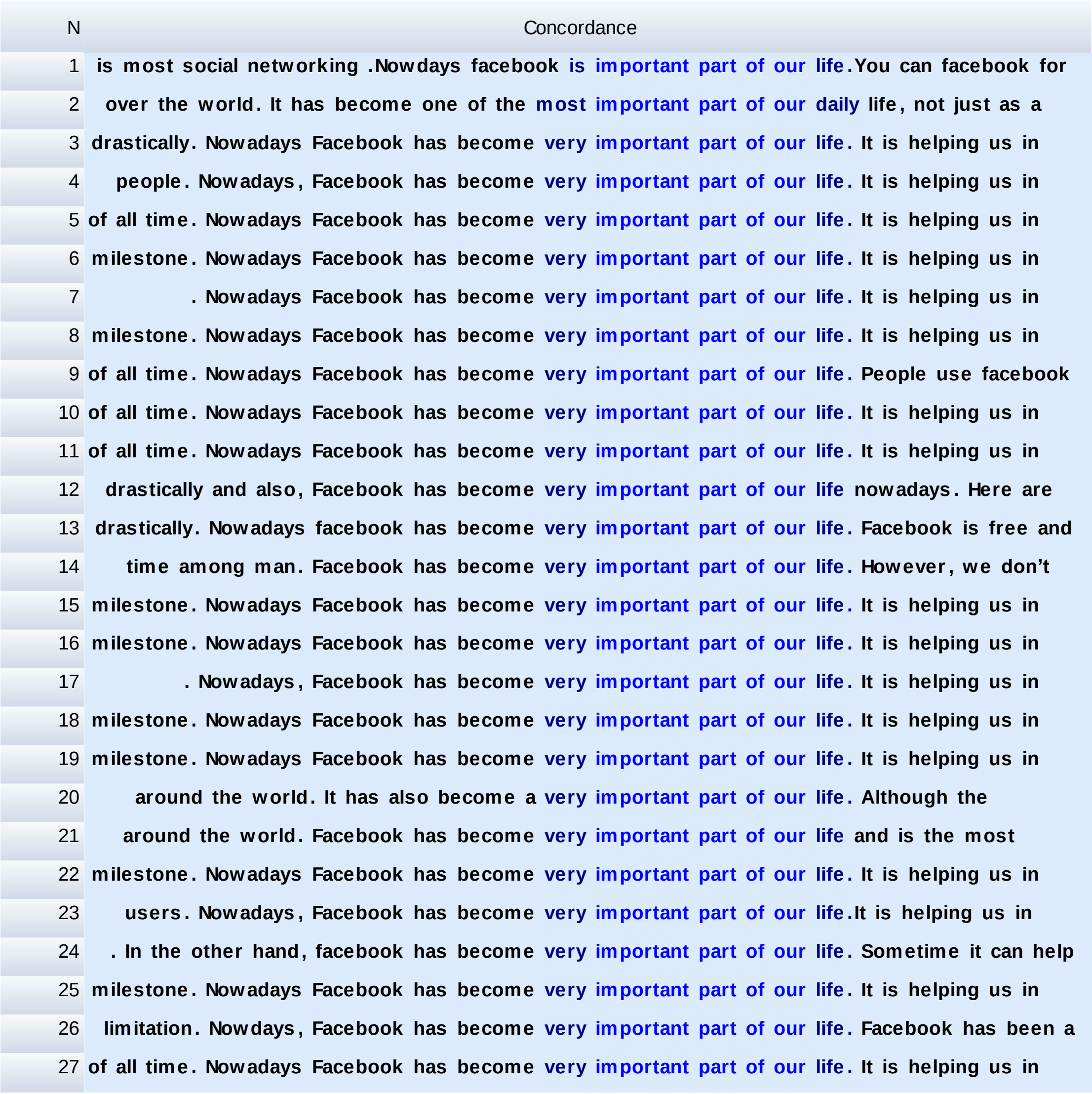
Figure 6.26: Concordance lines for *that we can* in LOCNESS

6.5.3.2.2 Importance

Two recurrent bundles that indicate importance, are *important part of our* and *become very important*. The bundle *important part of our* is found to be recurrent in 51 texts in MCSAW, and only once in LOCNESS: *Clothes are an* ***important part of our*** *every day lives and they always will be* (USARG.txt). Although all instances of this bundle are used to state the importance of the subject in focus, the bundle is clearly over-used in MCSAW.

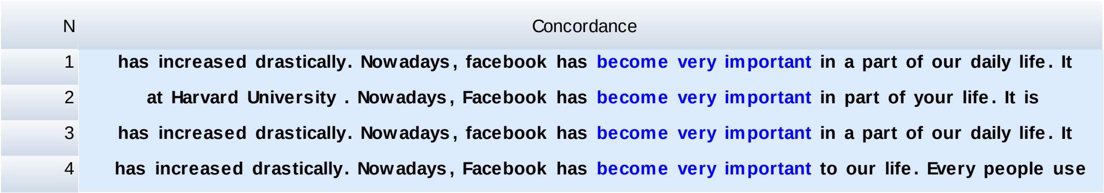
In much student writing, Flowerdew (2001: p. 367) mentions that the word *important* is used very frequently. This, she argues, is relatively similar to that which Granger and Tribble (1998) has revealed: learners were too reliant on superordinate adjectives such as *important* in their writing, which they used to the exclusion of words with a higher degree of specificity. However, the overuse of repeated lines *Nowadays Facebook has become very important part of our life* could, again, be argued as being the effect of essay prompt exercises in the writing classroom, or copying on the part of learners. In 52 occurrences of *become very important,* 48 are subsumed as part of the longer bundle *become very important part of our*. The remainder (4) are shown in Figure 6.28, presenting only variations of the same pattern rather than exact repetitions. Figure 6.27 shows 27 instances of the bundle occurring in MCSAW.

As can be seen, all 27 lines indicate the importance of Facebook as *part of our life**. Learner errors can also be detected as regards the singular aspect of *life* in *our lives* and the missing article before the bundle in *has become * very important part*. Use of the inclusive pronoun *our* not only reflects writer/reader visibility but, as discussed in the previous chapter, it also suggests learners' attainment for commonality or solidarity, i.e. strategy of shared experience between themselves as writers and their readers.



N	Concordance
1	is most social networking .Nowdays facebook is important part of our life.You can facebook for
2	over the world. It has become one of the most important part of our daily life, not just as a
3	drastically. Nowadays Facebook has become very important part of our life. It is helping us in
4	people. Nowadays, Facebook has become very important part of our life. It is helping us in
5	of all time. Nowadays Facebook has become very important part of our life. It is helping us in
6	milestone. Nowadays Facebook has become very important part of our life. It is helping us in
7	. Nowadays Facebook has become very important part of our life. It is helping us in
8	milestone. Nowadays Facebook has become very important part of our life. It is helping us in
9	of all time. Nowadays Facebook has become very important part of our life. People use facebook
10	of all time. Nowadays Facebook has become very important part of our life. It is helping us in
11	of all time. Nowadays Facebook has become very important part of our life. It is helping us in
12	drastically and also, Facebook has become very important part of our life nowadays. Here are
13	drastically. Nowadays facebook has become very important part of our life. Facebook is free and
14	time among man. Facebook has become very important part of our life. However, we don’t
15	milestone. Nowadays Facebook has become very important part of our life. It is helping us in
16	milestone. Nowadays Facebook has become very important part of our life. It is helping us in
17	. Nowadays, Facebook has become very important part of our life. It is helping us in
18	milestone. Nowadays Facebook has become very important part of our life. It is helping us in
19	milestone. Nowadays Facebook has become very important part of our life. It is helping us in
20	around the world. It has also become a very important part of our life. Although the
21	around the world. Facebook has become very important part of our life and is the most
22	milestone. Nowadays Facebook has become very important part of our life. It is helping us in
23	users. Nowadays, Facebook has become very important part of our life.It is helping us in
24	. In the other hand, facebook has become very important part of our life. Sometime it can help
25	milestone. Nowadays Facebook has become very important part of our life. It is helping us in
26	limitation. Nowdays, Facebook has become very important part of our life. Facebook has been a
27	of all time. Nowadays Facebook has become very important part of our life. It is helping us in

Figure 6.27: Concordance lines for *important part of our* in MCSAW



N	Concordance
1	has increased drastically. Nowadays, facebook has become very important in a part of our daily life. It
2	at Harvard University . Nowadays, Facebook has become very important in part of your life. It is
3	has increased drastically. Nowadays, facebook has become very important in a part of our daily life. It
4	has increased drastically. Nowadays, Facebook has become very important to our life. Every people use

Figure 6.28: Concordance lines for *become very important* in MCSAW

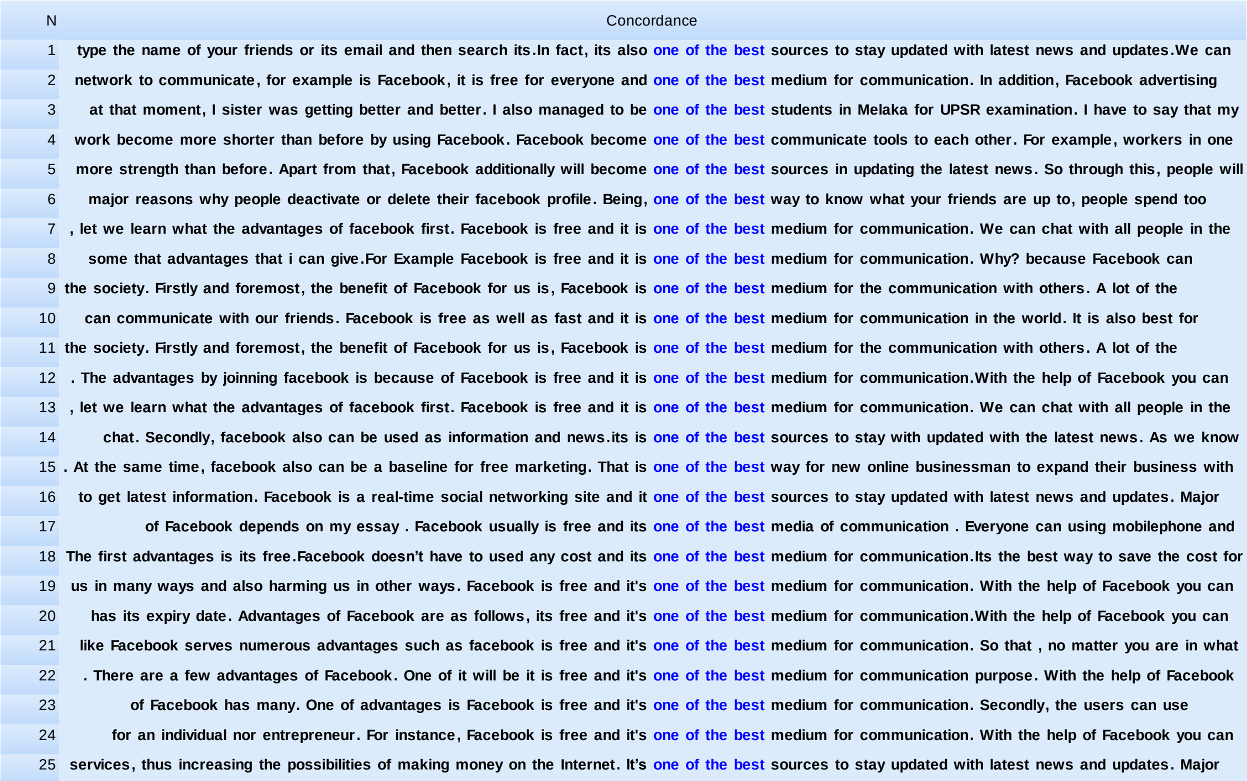
6.5.3.2.3 Emotivity

Two recurrent bundles reflect the expression of writer's evaluation of aspects of events as good or bad, in this case the expression of writer assessment of positive aspects (*one of the best* and *the advantages of*).

Figure 6.29 presents concordance lines for the recurrent bundle *one of the best* in MCSAW. Words co-occurring after *one of the best* include *sources*, *medium*, *students*, *communicate tools*, *way*, and *media*. Apart from *students*, the other collocates make reference to the topic Facebook. Mostly, this bundle indicates the writer's argument that Facebook is one of the best media for communication purposes. It is also found that the bundle is part of an independent clause (*it is/it's one of the best medium for communication*) that is connected to another independent clause preceding it (*Facebook is free*) by use of the conjunction *and*. In turn, the complex sentence emphasises the advantages or benefits of the social networking site through use of the lexis *free* and *best*. Although the recurrent 4-word bundle is found to be well-distributed in 69 texts in MCSAW, it only appears once in LOCNESS, as shown below. This indicates that the bundle is seldom used in the reference language variety. Another explanation may be the non-preference for the word 'best' in academic writing.

One of the best studies mentioned in Bergman's book answers this question: do students better understand scientific principles when taught from a two-model approach of origins (evolution and creation) or a one-model approach (only evolution or only creation)? (USARG.txt)

Furthermore, lines 7-13 and line 15 in Figure 6.29 show that *one of the best* incorporates the referential bundle *is one of the*, with lines 17-25 showing that this is also the case given the contraction *it's*. This adds to the over-used referential bundle *is one of the*, as already discussed in Section 6.5.1 above.

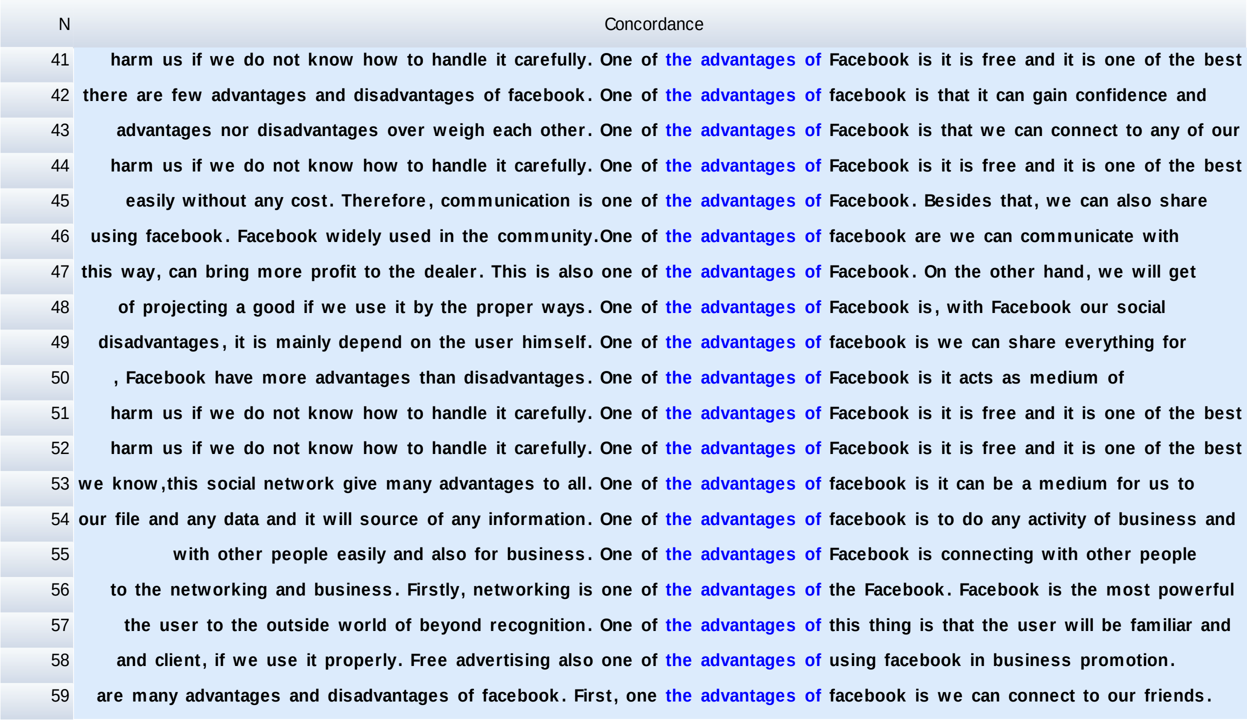


N	Concordance
1	type the name of your friends or its email and then search its.In fact, its also one of the best sources to stay updated with latest news and updates.We can
2	network to communicate, for example is Facebook, it is free for everyone and one of the best medium for communication. In addition, Facebook advertising
3	at that moment, I sister was getting better and better. I also managed to be one of the best students in Melaka for UPSR examination. I have to say that my
4	work become more shorter than before by using Facebook. Facebook become one of the best communicate tools to each other. For example, workers in one
5	more strength than before. Apart from that, Facebook additionally will become one of the best sources in updating the latest news. So through this, people will
6	major reasons why people deactivate or delete their facebook profile. Being, one of the best way to know what your friends are up to, people spend too
7	, let we learn what the advantages of facebook first. Facebook is free and it is one of the best medium for communication. We can chat with all people in the
8	some that advantages that i can give.For Example Facebook is free and it is one of the best medium for communication. Why? because Facebook can
9	the society. Firstly and foremost, the benefit of Facebook for us is, Facebook is one of the best medium for the communication with others. A lot of the
10	can communicate with our friends. Facebook is free as well as fast and it is one of the best medium for communication in the world. It is also best for
11	the society. Firstly and foremost, the benefit of Facebook for us is, Facebook is one of the best medium for the communication with others. A lot of the
12	. The advantages by joinning facebook is because of Facebook is free and it is one of the best medium for communication.With the help of Facebook you can
13	, let we learn what the advantages of facebook first. Facebook is free and it is one of the best medium for communication. We can chat with all people in the
14	chat. Secondly, facebook also can be used as information and news.its is one of the best sources to stay with updated with the latest news. As we know
15	. At the same time, facebook also can be a baseline for free marketing. That is one of the best way for new online businessman to expand their business with
16	to get latest information. Facebook is a real-time social networking site and it one of the best sources to stay updated with latest news and updates. Major
17	of Facebook depends on my essay . Facebook usually is free and its one of the best media of communication . Everyone can using mobilephone and
18	The first advantages is its free.Facebook doesn't have to used any cost and its one of the best medium for communication.Its the best way to save the cost for
19	us in many ways and also harming us in other ways. Facebook is free and it's one of the best medium for communication. With the help of Facebook you can
20	has its expiry date. Advantages of Facebook are as follows, its free and it's one of the best medium for communication.With the help of Facebook you can
21	like Facebook serves numerous advantages such as facebook is free and it's one of the best medium for communication. So that , no matter you are in what
22	. There are a few advantages of Facebook. One of it will be it is free and it's one of the best medium for communication purpose. With the help of Facebook
23	of Facebook has many. One of advantages is Facebook is free and it's one of the best medium for communication. Secondly, the users can use
24	for an individual nor entrepreneur. For instance, Facebook is free and it's one of the best medium for communication. With the help of Facebook you can
25	services, thus increasing the possibilities of making money on the Internet. It's one of the best sources to stay updated with latest news and updates. Major

Figure 6.29: Concordance lines for *one of the best* in MCSAW

The remaining recurrent bundle, *the advantages of*, is classified as attitudinal stance expressing emotivity because it expresses opinion through use of the word *advantages* in highlighting the positive aspects of the following head noun.

Out of 83 occurrences, 80 instances of this bundle are found to co-occur with *Facebook*, particularly *using/having Facebook*. This shows the writer's opinion about Facebook, in which writers advocate for the benefits of (using/having) Facebook. However, one instance is found to denote otherwise, *In a nut shell, the disadvantages of Facebook outweigh* <u>*the advantages of*</u> *Facebook and…*(MCSAW_1072.txt). This is because the bundle is found to co-occur with the verb *outweigh*, which signals the opposite: there are more disadvantages than advantages. Figure 6.30 presents some concordance lines for *the advantages of* in MCSAW. Interestingly, there is no occurrence of this bundle in LOCNESS. It is therefore, worth highlighting that prompt words *advantages* and/or *disadvantages* have been over-used in MCSAW, as has been discussed throughout the thesis so far, and thus implying that learners lack other ways to express the same meaning in their writing.



N	Concordance
41	harm us if we do not know how to handle it carefully. One of the advantages of Facebook is it is free and it is one of the best
42	there are few advantages and disadvantages of facebook. One of the advantages of facebook is that it can gain confidence and
43	advantages nor disadvantages over weigh each other. One of the advantages of Facebook is that we can connect to any of our
44	harm us if we do not know how to handle it carefully. One of the advantages of Facebook is it is free and it is one of the best
45	easily without any cost. Therefore, communication is one of the advantages of Facebook. Besides that, we can also share
46	using facebook. Facebook widely used in the community.One of the advantages of facebook are we can communicate with
47	this way, can bring more profit to the dealer. This is also one of the advantages of Facebook. On the other hand, we will get
48	of projecting a good if we use it by the proper ways. One of the advantages of Facebook is, with Facebook our social
49	disadvantages, it is mainly depend on the user himself. One of the advantages of facebook is we can share everything for
50	, Facebook have more advantages than disadvantages. One of the advantages of Facebook is it acts as medium of
51	harm us if we do not know how to handle it carefully. One of the advantages of Facebook is it is free and it is one of the best
52	harm us if we do not know how to handle it carefully. One of the advantages of Facebook is it is free and it is one of the best
53	we know ,this social network give many advantages to all. One of the advantages of facebook is it can be a medium for us to
54	our file and any data and it will source of any information. One of the advantages of facebook is to do any activity of business and
55	with other people easily and also for business. One of the advantages of Facebook is connecting with other people
56	to the networking and business. Firstly, networking is one of the advantages of the Facebook. Facebook is the most powerful
57	the user to the outside world of beyond recognition. One of the advantages of this thing is that the user will be familiar and
58	and client, if we use it properly. Free advertising also one of the advantages of using facebook in business promotion.
59	are many advantages and disadvantages of facebook. First, one the advantages of facebook is we can connect to our friends.

Figure 6.30: Concordance lines for t*he advantages of* in MCSAW

6.6 Summary

This chapter contrasted and examined learner writing in MCSAW with respect to the reference language variety, i.e. LOCNESS in terms of shared 4-word and 3-word lexical bundles identified by WordSmith Tools. Given the overall frequency distribution of lexical bundles and their qualitative analysis in concordance lines, it can be said that there are certain bundles that are shared in both the Malaysian learner corpus and reference language variety, LOCNESS. Key similarities in the writing of MCSAW and LOCNESS writers include using bundles such as *is one of the*, *most of the* and *in my opinion* for similar functions. However, other bundles that were shared between the corpora highlight key differences in the writing of the two groups of novice writers (e.g. *the popularity of*, *people around the world*, *anywhere in the world*, *first of all*, *this is because*).

Findings indicate that the shared 4-word bundles are mostly referential in MCSAW, with no discourse-organising bundles occurring in the list. This leads us to believe that the highly significant 4-word bundles in learner writing are mainly used to refer to physical, abstract or contextual aspects, including those that focus on a particular feature of an entity as important. These bundles are, hence, mostly topic dependent. Frequent bundles were also found to co-occur with high-frequency words such as *big*. Furthermore, the recurrent bundle *people around the world* was found to co-occur with the predeterminer *all*, which results in bundles such as *all people around the world* and *all the people around the world*. In the reference language variety, however, this bundle occurred only twice, in two varieties: following the *-of* construction (*millions of people around the world*), and continued by a relative clause (*all the people around the world who enjoy watching the sport*…). It was also found that certain bundles appear to structure the text topically, providing cohesion, and as a result, function as a 'hyper-Theme' (e.g. *many ways* are specified in what follows, notably by words *like*, *either* and the following sentences).

On the contrary, shared 3-word bundles in MCSAW were identified as a mixture of Referential, Discourse-organising, and Stance bundles. In fact, 3-word bundles were found three times more than 4-word bundles, which is not surprising since shorter bundles tend to be more frequent than longer bundles. Shared 3-word bundles include discourse organisers such as *first of all* and *this is because*, that may suggest learners' underuse of longer string of discourse organisers. The shared 3-word bundles also indicated that learners frequently use the personal pronoun *we* (i.e. *that we can*) and the modal verb *can* (i.e. *can be a*), including combinations of *we can* bundles such as *we can make*, *we can use*, and *we can know*. These instances relate to the highly significant keywords *we* and *can* as discussed in the previous two chapters, and in turn relate to both the speaker/writer and reader, which is possibly intended to appeal to the strategy of shared experience as discussed in Chapter 5. However, these bundles were found to be highly personal and spoken-like, which points to learners' writing as being more assertive and less tentative than that found in the reference language variety. More specifically, learners frequently use attitudinal stance bundles, especially with personal pronouns (e.g. *we want to*, *we have to*), resulting in highly interpersonal writing. Arguably, this could be a result of the topic or genre of argumentative writing, which encourage writers to convince their audience, i.e. through the expression of personal opinions. Such bundles were mostly found to function in expressing ability (*us the opportunity to*), importance (*important part of our*; *become very important*), and emotivity (e.g. *one of the best*; *the advantages of*).

Results also revealed that certain bundles are likely the result of repeated sentences (prompts, templates) rather than indicative of learner style: for instance, bundles that are underlined in the following sentences, *With the help of Facebook, we/you can connect to different people from <u>anywhere in the world</u> because almost every people around the world use*

Facebook; *But now Facebook gives us the opportunity to communicate with our old friend very easily without any cost*; and *Nowadays Facebook has become very important part of our life*. In fact, some repeated bundles may be part of longer recurrent bundles (e.g. *us the opportunity to communicate with*; *one of the best medium for communication*; *one of the advantages of Facebook*), and are clearly topic-dependent. Similar findings show instances of redundancy or tautology in the use of bundles, in that repeated lines were found in more than one essay and thus, do not reveal much about learners' language. Rather, this suggests plagiarism or the possible influence of prompts that were used in the classroom.

Nevertheless, other bundles proved to corroborate previous studies that claim they are descriptive of learner style. These instances include deictic bundles, especially ones referring to the world (*people around the world*; *anywhere in the world*), the identification bundle *is one of the,* and attitudinal stance bundle expressing importance (*important part of our*; *become very important*). Moreover, prompt words such as *advantages* and *disadvantages* were also found to be problematic, which suggest learners' limited vocabulary repertoire and controlled style in writing. These observations share similar findings to those in previous studies on lexical bundles, which found an overuse of particular lexical bundles (Cobb, 2003; De Cock, 1998). Further qualitative analyses revealed that bundles with the words *advantages*/*disadvantages* (e.g. *the advantages of*) express emotivity, as it expresses opinion - highlighting the positive aspects of the following head noun. Other recurrent stance bundles referred to attitudinal expressions such as *us the opportunity to*, *important part of our*, *become very important*, and *one of the best*. Although bundles were found to be shared in at least one occurrence in the reference corpus, their occurrences in MCSAW appear somewhat odd, and even repetitive, because of the overuse of the limited and less varied bundles that learners possess.

Qualitative results also showed that transfer effects are noticeable in the overuse of lexical bundles whose equivalent forms fulfil specific discourse functions in Malay, i.e. *is one of the* (*merupakan salah satu daripada*) *this is because* (*ini disebabkan/ini kerana*) and 'we can' (*kita boleh*) bundles. In addition, results also conclude that the types of bundles learners produce are less varied when compared to their native speaker counterparts. Although quantitative results show a majority of 4-word and 3-word bundles comprising verb phrases, qualitative results reveal that most of these bundles were used to signal referential meaning. This means that not only do native speakers have a broader repertoire of bundle types, but they also tend to display greater variety in form. It could be that certain groups of recurrent bundles are under-used by learners, while others are found to be over-used (i.e. bundles that are tautological). Overuse or underuse of items in writing could be explained by lack of knowledge in the discourse functions of words, thus highlighting learners' limited vocabulary repertoire. Therefore, this chapter revealed 4-word and 3-word lexical bundles that are typical of Malaysian learner writing, which could have pedagogical implications. For instance, Tribble (1991) suggests that examples of concordance data from a learner corpus could be exploited in the classroom by having learners work on re-wording the concordance lines and encouraging them to use a broader range of vocabulary.

One last important note to be made is the limited topics provided in MCSAW. Although there are numerous learner corpus studies that have shown meaningful insights by using a variety of types of specialised corpus (e.g. Imm, 2009; Biber & Barbieri, 2007; Flowerdew, 2001), this chapter revealed that there are definitely limitations as regards the overuse of topic-related words and repetition of essay prompts, which were found to be prevalent in learner writing. It has been found that there is greater variation of nouns in the reference language variety compared to the learner corpus, thus indicating topic variability, which is limited in MCSAW. As a result, it is important to take caution in using 'range' analysis for down-sampling purposes, and that this may not work that well for corpora with limited topics and/or use of repeated phrases that are the result of prompts, templates or other classroom teaching. However, the exploration of different bundle lengths and method of distribution analysis alongside qualitative examination of concordancing have potentially identified bundles that are characteristic of Malaysian learners, albeit a constant display of words that are specifically topic dependent given the respective text type. Malaysian learners, thus, often find their use of bundles problematic, typically over-using a limited number of well-known phrases, while at the same time lacking a diverse enough phrasal repertoire to employ lexical bundles in a native-like manner. In another sense, it is only by a single topic corpus that the limitations in vocabulary can be established.

Chapter 7: Conclusion

7.1 Introduction

This thesis investigated Malaysian learner argumentative writing via a contrastive, corpus-driven approach. By comparing and contrasting the Malaysian Corpus of Argumentative Writing (MCSAW) against the Louvain Corpus of Native English Essays (LOCNESS), the study presented new findings on Malaysian learner language, through employing the CIA approach. The ensuing sub-sections include an overview of the significant findings, followed by a discussion of limitations, recommendations for future research, and implications for ESL pedagogy.

7.2 Overview of significant findings

Chapter 4 firstly identified and discussed modality as one salient feature of learner writing in MCSAW. More specifically, the modal verb *can* is, statistically, a highly significant keyword, occurring across 97% (495 out of 509 texts) of the Malaysian corpus. Observations of *can* against the demographic background of MCSAW writers suggest that it is widely used among all three major L1 groups of learners in the corpus, suggesting two possible explanations: either this occurrence is not purely indicative of learners' first language (L1) influence (but rather influenced by another factor altogether), or they are all equally influenced by their L1. These findings concur with past research on the highly frequent use of modal verb *can* in Malaysian learner writing (Mohamed Ismail et al., 2013; Mukundan, et al., 2013). However, it must be noted that there is an uneven distribution of MCSAW writers with different L1 backgrounds, and essay topics are limited to two. Collecting more argumentative essays on a variety of different topics, and ensuring that L1 groups are included to the same extent, could thus offer further insights.

In addition to being a keyword with a very high range, it was also found that *can* is spread within MCSAW texts in a uniform dispersion plot. This means that learners use *can* in all parts of their essays, namely in introducing the proposition of argument, discussing the argument, and synthesising the discussion as well as affirming the validity of the proposition. In general, the frequent use of this modal verb points to the persuasive genre of argumentative writing, as well as to the essay prompts (as the suggestion to discuss advantages/disadvantages leads to a high use of the modal verb *can* in this type of persuasive essay).

Furthermore, collocational analysis of *can* reveals common lexical and grammatical patterns of the modal verb's co-occurrence in MCSAW. It was found that certain patterns of language are used similarly in both corpora, such as the use of *we can* phrases with high frequency verbs such as *make*, *find*, and *enjoy*. These patterns are found to be characteristic of spoken discourse and are uncommon in academic texts (Granger & Paquot, 2009; Lee & Chen, 2009). It is likely, then, that these are characteristic of novice writers, since they both occur in MCSAW and LOCNESS, thus also highlighting the benefits of comparing MCSAW with another novice group of writers. In contrast, collocates *from*, *so*, *later*, and *found* were used differently between MCSAW and LOCNESS. Essentially, this highlights several features of learner language in MCSAW, including traces of possible L1 transfer (e.g. *also can*), features that are more characteristic of everyday talk than written language (e.g. *So, we can say*), and the repetition of similar sentences across texts in MCSAW. For example, in the investigation of the collocate *found*, the sentence *From different sources it is found that, Facebook can be life threatening sometimes* occurs more than one time, although from different texts. This could be evidence for plagiarism within texts, or the overuse of certain prompt sentences taught in the classroom or templates provided prior to the essay production. Recurrences of tautology in learner texts, which are seen throughout all chapter analyses, present serious issues, especially as regard the use of this corpus in Malaysian LCR (as further discussed below).

Moving on, qualitative analysis (concordancing) shows that the modal verb is mostly used to function as the 'Ability' meaning (79%), followed by 'Possibility' (16%) and 'Permission' (1%). This initial observation supports past findings from Mohamed Ismail et al. (2013) that *can* is mostly used in MCSAW to express a sense of ability rather than other functions of modality. More specifically, *can* denoting ability is expressed for both animate (e.g. *users can*…) and in-animate subjects (e.g. *Facebook can*…). More importantly, results indicate that *can* is mostly used in relation to its use in the first language. For example, *can* expressing 'Ability' is mostly found within *can get* phrases, similar to how it would be translated in the Malay language – *boleh dapat.* Similarly, the phrase *we can* also highlights the 'Ability' meaning in the translated version of *kita boleh*, thus indicating the possibility of L1 transfer. Effects of the L1 are also evident in *can* denoting the 'Possibility' meaning, especially in combinations of *that can* phrases that resemble the Malay equivalent *yang dapat*. These examples, which indicate potential L1 influence, are argued to be due to 'Possibility' meanings of *can* that are related to some form of opportunity, benefit, or advantage. Therefore, its use in the highly frequent *can* constructions show how *can* is used in such a way as to denote some sense or relation to the topic and/or argumentative genre, which were not discussed in past Malaysian LCR.

In sum, analysis of *can* is argued to be pervasive in MCSAW for three main reasons – the modal verb is not only found to be statistically significant, it is widely used among MCSAW writers as well as spread widely within texts. Further investigation in both corpora reveals that both groups of novice writers produce similar patterns of *can* (*we can* + high-frequency verbs), which are indicative of spoken language. However, different patterns of *can* in MCSAW highlight distinctive patterns that may be idiosyncratic to MCSAW writers. These include several prepositional phrases and lack of passive structures in learner writing. Two major explanations for these observations could be the topic/genre of argumentative writing and effects of L1 transfer. In fact, further qualitative analysis reveals that the prevalent use of *can* to denote the ability/possibility meanings is also linked to the influence of Malay and essay type.

In addition to the modal verb *can*, the thesis identified another salient keyword in MCSAW, the personal plural pronoun *we*, as also being statistically significant, and widespread across 84% (429 out of 509 texts) of the Malaysian corpus. It is also found that *we* is frequently used among all writers in MCSAW, irrespective of L1 background. Thus, results for use of pronouns have supported past research (Cobb, 2003; McCrostie, 2008), in which learners of English produce higher occurrences of the first person pronouns than in the reference language variety. In terms of the dispersion plot, Malaysian learners use *we* more in the middle and end of their writing than in the beginning. This suggests that the pronoun may have a role to play in particular discourse strategies, which was confirmed by collocation and concordance analysis as explained below.

Collocation analysis revealed that both groups of novice writers (MCSAW and LOCNESS) appear to use *we* with words expressing modality and verbs of necessity/desire (*need, must*), mental verbs and verbs of discovery/perception (*know, find, see*), action and speech verbs (*live, ask, say*), reflexive pronoun (*ourselves*), and adverbs (*today, already*). More generally, results of the collocational analysis indicate that MCSAW and LOCNESS writers use *we* to engage with their readers, in what is called the 'solidarity strategy' (Harwood, 2005b;

Hyland, 2002a; Luzón, 2009). Findings in Chapter 5 indicate that this is a major persuasive strategy used by Malaysian learners in these types of argumentative essays. In addition, phrases that include both *we* and *can* (e.g. *we can find*, *we can see,* and *we can ask*) appear to be used by the MCSAW writers as a type of subjective and 'collective' hedge, instead of more impersonal constructions. Interestingly, results from collocation comparison of *we* indicate similar use of the personal plural pronoun among both groups of novice writers. This suggests similar strategies of using *we* in writing argumentative essays or the comparable level of novice writing, in which first person pronouns (*we*) are mainly used to signal interpersonal discourse and direct involvement of the writer (Hinkel, 2002). In short, results in Chapter 5 indicate that both MCSAW and LOCNESS writers exhibit a high amount of writer/reader visibility in their argumentative writing, demonstrating that the investigation of Malaysian learner writing against another novice group of writers reveals insights into novice writing. This also shows the value of comparing and contrasting both varieties, rather than treating the L1 variety as the 'norm'.

The concordance analysis of the functions of *we* in MCSAW provided further detail. It suggested that the personal pronoun is used mostly in assuming shared experiences/knowledge, goals, beliefs (86.3%); this is followed by expression of opinion or volition (8.9%), stating conclusions (1.6%), calling the reader's attention (1.3%), and in guiding the reader through the text as well as stating purposes (both 1.0%). This confirms the assumption that MCSAW writers use *we* to associate themselves and share experiences with readers, achieving 'solidarity'. Furthermore, in expressing opinions, learners combine *we* with verbs of necessity (*we need to*) and deontic modality (*we should*, *we have to*). As repeatedly mentioned, this may be a feature of the argumentative genre in which writers are encouraged to be persuasive, and therefore, opinions are usually expected. However, opinions voiced by learners are often found to be influenced by their LI, for example, the phrase *we just need to* (*kita hanya perlukan*), and *we actually/definitely* (*kita sebenarnya/sememangnya*), as shown in Chapter 5.

Finally, Chapter 6 has explored recurrent lexical bundles that are prevalent in learner writing. More specifically, statistically significant 3- and 4-word sequences were analysed with at least one occurrence in the reference language variety. Out of the 26 4-word shared bundles, only 7 (26.9%) occurred more than twice in the reference corpus. On the other hand, 26 (41.3%) of the 63 shared 3-word bundles were found occurring more than twice in LOCNESS. This means that salient 4-word and 3-word bundles were not significantly shared in LOCNESS.

Structurally, both shared 3- and 4-word bundles constitute verb phrases in contrast to noun and prepositional phrases, the latter which were said to be more characteristic of academic prose (Biber, 2009). Frequent 'be' constructions were also found, similar to findings for expository and argumentative essays by other ESL learners (Chen & Baker, 2014). Notably, there were frequent recurrences of personal pronouns (e.g. *us the opportunity to*, *they are too*), modality (e.g. *can help us to*, *can make us*), and copula *be* constructions (e.g. *is one of the*, *it is easier*) within the shared bundles. In fact, combination of both pronouns and modal verb *can* in a single bundle were also found (e.g. *that we can*, *we can find*), thus further justifying the overuse of *can* and *we* in Chapter 4 and Chapter 5 analyses, respectively.

As regard the examination of shared bundles according to discourse functions, overall most over-used bundles in MCSAW were identified as 'Referential' (14 4-word bundles, 28 3-word bundles), and are mostly topic-oriented. Similar to past findings (Ädel & Erman, 2012; Chen & Baker, 2014), referential bundles *is one of the* and *there are some*, which are more typical of conversation than writing (Biber et al., 2004; Chen & Baker, 2014), were found highly frequent in learner writing. 'Stance' bundles were less frequent (12 4-word bundles, 26 3-word bundles), and when they do occur they are mostly topic-related. Some stance bundles include the prompt words *advantages* and *disadvantages*, while others include superlative 'best' such as *the best way to*, and words indicating volition such as *you want to*, which demonstrate writers' direct and strong assertion of their argument/opinion. Interestingly, 'Discourse organisers' were not found within 4-word bundles, but exist in a number of 3-word discourse organisers such as *first of all*, *for example if*, and *as a conclusion*. The investigation of different bundle lengths proved to be beneficial as results indicated that 4-word bundles are less varied compared to 3-word bundles in MCSAW. One explanation would be that learners may not know enough long bundles to be able to use them in writing. Qualitative results also showed that transfer effects are noticeable in the overuse of lexical bundles, which are used similarly in Malay, i.e. *is one of the* ('merupakan salah satu daripada'), *this is because* ('ini disebabkan/ini kerana'), and *we can* ('kita boleh') bundles.

In general, all three chapters have identified a significant number of cases where it is likely that language use was affected by the topic or essay prompt. In other words, a number of findings are perhaps not indicative of learner style/learner writing in general, but instead are more likely to be affected by the topic or essay prompt or the result of repetition across essays. Other patterns were found to be repeated due to effects of L1 transfer, as noted above.

Interestingly, results have shown that even a grammatical word such as *can* may clearly be influenced by the essay topic/prompt, although grammatical words are usually considered as indicators of style rather than aboutness (Baker, 2004). This ultimately shows that "the usefulness of a learner corpus is directly proportional to the care that has been exerted in designing it and compromising the design stage inevitably leads to less solid results" (Granger, 2008: p. 338). Hence, this thesis also highlights limitations of using MCSAW in future LCR.

To summarise, this thesis has employed the CIA framework via a corpus-driven approach, particularly through consideration of comparable corpora, paying close attention to examining range and dispersion, as well as reflecting on the topic/genre of argumentative writing overall. This resulted in a rich number of empirical findings. As has been discussed, Malaysian learners have significantly shown an overuse of the modal verb *can*, plural pronoun *we* and referential-type 3-word bundles and 4-word bundles in comparison to their novice native-speaking counterparts in argumentative writing. Most importantly, while much past research has identified learner writing as exhibiting speech written down (Gilquin & Paquot, 2008; Paquot, Hasselgård & Ebeling, 2013), few have explored external factors (e.g. essay topic and genre, L1 transfer) behind such accounts. In the following sections, a series of reflections will be presented in terms of the pros and cons of using the current methodology, recommendations for future research, and implications for ESL pedagogy.

7.3 Contributions of the study

As Granger (2015) restates, CIA has been a highly popular method in LCR for more than twenty years. More specifically, comparing learner corpora against a suitable reference language variety has been shown advantageous in past scholarship (e.g. Gilquin, 2001; Xiao et al., 2006). These benefits can be appraised according to several areas, namely varieties of the languages investigated, medium and genre of discourse, proficiency level, linguistic phenomena explored, and type of CIA approach. In this section of the chapter, I will evaluate the present study based on the aforementioned issues, as well as highlighting some limitations at the end of this section.

Firstly, with many CIA studies, English has been the preferred target language to investigate. However, the present study focuses on a type of L2 English: the Malaysian learner English variety, which has not been the subject of much investigation in LCR so far. In response to the notion of varieties promoted through the newly revised CIA (Granger, 2015), by investigating MCSAW, we are able to explore and uncover linguistic features that characterize the interlanguage of Malaysian second language learners, particularly at the advanced (college) level, via empirical data. In addition, contrasting MCSAW against the comparable reference language variety, LOCNESS, reveals features of learner language that are specific to the text type, i.e. argumentative essay and novice writing.

In terms of proficiency level, MCSAW college texts were chosen, instead of essays written by 16- and 17-year old students, in order to reach near-comparability with LOCNESS texts that were all written by A-level or/and college students. In this regard, MCSAW language users are considered to be in the advanced stage of interlanguage (Granger, 2015). It is also important to bear in mind that both sets of essays were written by novice writers, and therefore resemblances are found in certain uses of lexical items due to similar strategies adopted by novice writers in writing argumentative essays. This is where CIA is beneficial: not only is it possible to examine features that are particularly more significant of Malaysian learners' writing, but also how similar structures could point to particular discourse functions of the argumentative-type essay. However, it would be interesting to examine and compare different proficiency levels of writers within MCSAW texts in future research, to evaluate language development across age.

As regard what/which linguistic phenomena are investigated, the present thesis has shown that, by examining data from the bottom up (starting with keywords analysis and range), statistically significant and well-distributed items are selected without pre-judgements (Lee & Chen, 2009). Ultimately, one can be sure that Malaysian writers of MCSAW over-use the modal verb *can* and personal pronoun *we* in their argumentative writing. Analysing the occurrence of items across a number of texts (range) contributes to the innovative means taken in the present study, in contrast to past LCR studies that have ignored whether an item is widely used by different learners (i.e. considering only frequency but not range) (Gilquin et al., 2007). By investigating range, one is more confident that the item in question is not only statistically more frequent in the learner corpus, but is also widely used among learners. In other words, examining recurrence of items (both individual and lexical bundles) in multiple texts suggests at least some perceptual salience among users, and thus a particular writing style (Hyland, 2012). However, this thesis has also shown weaknesses in using range with respect to a corpus such as MCSAW, where only two topics are present, since some of the results appear linked to the topic/essay prompt. As already noted, other results derive from the use of identical phrases by students across texts (the likely effect of copying or classroom teaching). As a result, range may not work that well as a down-sampling technique where a corpus such as MCSAW is concerned. Nevertheless, using this technique in the present thesis has allowed for this issue to be identified.

Moreover, this thesis has further extended the analysis by investigating both individual and lexical bundles, in contrast to most studies in LCR, which investigate individual and multi-word items separately (Bestgen & Granger, 2014: pp. 233-234; Crompton, 2005: pp. 159-160). Apart from the frequently used referential-type bundles, it can be argued that results of lexical bundles in Chapter 6 confirm that the modal verb *can* and personal plural pronoun *we* in Chapters 4 and 5 are indeed over-used and prevalent in longer strings of words. This means that one can arrive at stronger claims that *can* and *we* are characteristic of the Malaysian L2 English phrasicon,[68] more specifically in argumentative texts. For classifying bundles, this thesis used a new categorisation scheme necessary for identified bundles to be classified more precisely. This is just one example of several cases in this thesis where I adapted classification schemes in innovative ways to allow for a comprehensive analysis and interpretation of the data, as well as to take into account the specific genre of writing. In so doing, this thesis also makes an original contribution to the development of analytical frameworks in LCR.

In terms of computerised analyses, the thesis demonstrated a number of features that prove to be powerful in LCR. This includes the use/testing of the range function as a down-sampling method, as already discussed above. In addition, using the plot function of WordSmith Tools has allowed for a small amount of intra-textual analysis, often lacking in corpus linguistic research (Flowerdew, 2005: p. 329).

In addition, using concordancing for the qualitative analyses of the discourse functions of salient items challenges criticisms about learner corpus analysis that only allow interpretation of descriptive statistics (Gilquin & Granger, 2015). As McEnery and Hardie (2012: p. 176) state, “the joint quantitative–qualitative analysis typical of corpus linguistics lends itself very readily to the study of the functional-formal links”, which has been exemplified in the analyses of *can, we* and lexical bundles of the present thesis. Using the concordance tool also allowed me to consider potential L1 transfer and the impact of the topic/genre on the choice of lexis and functions in which items were frequently used (Lee & Chen, 2009). In this way, this thesis

[68] I.e. “the whole set of formulaic sequences in learner language” (Paquot & Granger, 2012: p131)

has made a substantial contribution to studies in LCR that combine quantitative and qualitative analysis.[69]

Although the thesis demonstrated that the contrastive corpus-driven approach is productive in investigating Malaysian learner argumentative writing, it is not without weaknesses. Several limitations are duly noted, which should be borne in mind when interpreting the findings. Firstly, it is important to make note that, whatever is found with regard to the exploration of MCSAW versus LOCNESS, the findings cannot be generalised beyond the scope of the writing genre and the group of novice writers of the corpora. This means that results only account for the specific text type, written by the advanced (college) level group of learners in comparison with A-level/university students in the reference language variety. MCSAW is also relatively limited in size and scope compared to some other corpus studies. Statements about learners' competence on the basis of performance data in MCSAW, in turn, must remain speculative: that is, results cannot be generalised to all ESL users in Malaysia. This is also because MCSAW consists of different L1 groups of writers. Furthermore, even though essays are comparable in terms of their argumentative genre, data was collected in entirely different contexts and the essay topics written by both groups of novice writers were not exactly the same. In addition, the thesis only included some qualitative contrastive analyses of *we* and *can.* Future research needs to determine if particular types of modality are over-used. In addition, lexical bundles were investigated as limited to the span of five words to the left and right of node word, as well as looking at 3- and 4- word length bundles only. Findings for other bundle lengths and span would, in turn, show different results. It is also important to remember that under-used key bundles were not analysed, which represents an important area for future research.

Clearly, the present study follows the L1 vs L2 type of CIA analysis (as discussed in Chapter 2). This means that the method of comparing and contrasting language used in MCSAW and LOCNESS is between English first language users (LOCNESS) and English second language users (MCSAW). As many researchers have demonstrated (Gries & Divjak, 2009; Xiao et al., 2006), learner language is best investigated via comparison with a reference (specifically a native speaker) corpus. Not only is LOCNESS comparable in terms of the argumentative genre, writers of the reference language variety are also considered to be novice

[69] "Concordances and frequency data exemplify respectively the two forms of analysis, namely qualitative and quantitative, that are equally important to corpus linguistics" (McEnery & Hardie, 2012: p2).

writers and thus, in turn, appropriate for the type of CIA conducted in the thesis. However, this thesis only provides the interlanguage (IL) of one particular group of Malaysian writers, and therefore further contrastive interlanguage-type studies (L2 vs L2) are encouraged, providing comparisons to be made between different age groups of Malaysian learners in order to investigate their development in acquiring English as a second language. As Granger (2015: p. 20) notes, it is important for CIA "to extend our model beyond interlanguage varieties".

More importantly, this thesis continuously reports many repeated words, phrases or structures within the learner corpus, exemplified across numerous texts, and therefore warrants further investigation for the use of this specific corpus in CIA studies. Several results are clearly linked to the essay topic of Facebook, rather than being indicative of learner style more generally. One way to explain this is in the limited essay topics in MCSAW. While the compilers state that one of the reasons for this was to have a familiar, generic topic to encourage learners to better write (Mukundan & Kalajahi, 2013), I would propose that future Malaysian written corpora comprise more than two essay topics, to avoid this limitation. Further examination of essay topic effects (e.g. Hinkel, 2009; Huat, 2003) would also appear to be necessary.

Nevertheless, this thesis aimed to follow the design and methods of conducting the contrastive corpus-driven approach. These include performing automatic analyses such as the keywords analysis as a starting point for investigation, conducting (more) manual analysis of the learner corpus such as investigating discourse functions relative to past scholarship, and adapting classifications to suit the genre of writing, as well as restricting analysis to certain combinations or words through examining the distribution of items across a number of texts (i.e. range). In spite of the limitations, this study has shown the significance of using corpus methods in investigating Malaysian learner English writing. More specifically, this thesis has presented salient features of Malaysian learner English writing that could potentially be useful in understanding learners' interlanguage. In addition, the present study presents potential areas for the improvement of Malaysian learner corpora in the future. As Marchi (2013: p. 101) puts it: "any analysis is just a snapshot of some point in its life. And yet, […] the analysis may nevertheless be a brush-stroke adding to the big picture, its results may find resonance in other findings and be used as input for other studies".

7.4 Recommendations for future research

Learner corpora are a fairly recent phenomenon, as they only started to emerge in the 1990s, more than 30 years after native speaker corpora began to be compiled (Nesselhauf, 2004: p. 127). Thus, although there is currently much activity in the field, most existing learner corpora are incomplete, and studies based on learner corpora are only starting to become more widespread, especially in the Malaysian context. Given more research in the field, particularly studies employing CIA, would lead to better-informed and improved LCR studies in Malaysia.

As many linguists have stated (Lee, 2008; Römer, 2006), knowing what types of corpora are out there enables one to better suit the data to one's research objectives. While comparability issues are hard to dismiss, further research that involves building a specialised corpus pertaining to detailed and consistent criteria may produce significant results. For instance, in examining differences against a reference language variety such as LOCNESS, it would be purposeful to compile similar texts, that try to match similar contexts, writing setting, and even essay topics, in order to reach near-comparability between corpora. Similar to existing corpora such as the International Corpus of Learner English (ICLE) and International Corpus of English (ICE), future Malaysian learner corpora should adhere to firm guidelines for collecting data that would enable it to be compared easily with another reference corpus in the future. Collaboration between Malaysian corpus-building teams should also be encouraged to facilitate comparison between different varieties of language use in Malaysia.

CIA studies focusing on languages other than English are also much encouraged (Granger, 2015); and hence, CIA studies that investigate and compare the Malay language with another reference language variety would be insightful, considering the many instances of L1 transfer as shown in this thesis. Similarly, it would be beneficial to investigate more than one interlanguage variety (i.e. two or more types of learner corpora), for example, in the exploration of English language varieties (e.g. Nesselhauf, 2009). This type of CIA study also enables diachronic corpus research to be undertaken in examining learners' proficiency across time. Such studies (i.e. investigating different sets of learner texts across time) have been shown to be fruitful in describing language change (e.g. Xu & Liang, 2012), and are thus valuable in the development of Malaysian LCR.

As regard the study of modality in learner writing, it could be useful to investigate different word-forms of modal verbs such as the negative form for *can*, i.e. *cannot* (*can't*). It would also be interesting to examine whether cultural or regional factors encourage the use of writer-visibility, i.e. do Malaysian learners produce more inclusive *we* due to the Eastern culture, or does the prevalence of high personal pronoun use vary according to learners' regional background? This can be done by conducting another CIA study, exploring more than one type of learner interlanguage (learner corpus), specifically within the South-East Asian region. In addition, examination of different bundle lengths across a number of diachronic corpora could provide further insight into learners' phraseological development.

Genre analysis is another significant approach to LCR. Given the textual analyses presented in corpus linguistics studies, it would be worthwhile to complement the corpus-driven approach with a genre analysis (Flowerdew, 2005). As a result, it would be interesting to see whether Malaysian learners internalise their knowledge of writing argumentative-type essays differently than for other genres of writing.

7.5 Implications for ESL pedagogy

Similar to other studies in LCR, this thesis not only hopes to contribute to the body of knowledge in the area, but also suggests ways in which results of the present study can be beneficial for those wishing to develop ESL pedagogy, particularly in Malaysia's ELT classrooms. In other words, this sub-section aims to answer the questions, 'What have we learnt from this study?', and 'How does it help Malaysian classrooms in particular?' This thesis presents three key implications for the specific field, while taking note of the general implications of using corpora in the ESL classroom.

Salient features such as the highly frequent use of *can* and *we* in MCSAW illustrate learners' writing style, specifically in the argumentative-type essay, which asks students to consider advantages and disadvantages of entities, behaviour, etc. Although tendencies only point towards the Malaysian learner English variety, teachers may want to make use of this information, such as encouraging students to vary their use of modality, and/or teaching students how to produce more impersonal statements in their writing. In addition to studies that have identified the over-teaching of modal verb *can* in Malaysian English textbooks (Mukundan & Khojasteh, 2011), results from the present study add to the conviction that other modal verbs should be given similar weight in the teaching of modality. This is also the case with the pervasive *we* in learner writing (McCrostie, 2008), which shows high writer/reader visibility in MCSAW. In turn, students should be encouraged to reduce overt author presence in their writing in order to avoid their writing as resembling speech written down (Gilquin & Paquot, 2008). Another feature of learner language in MCSAW is the highly frequent use of referential-type bundles, at the expense of other types of bundles, especially discourse-organisers. As noted in past research (Ebeling, 2011: p. 66), "English essays are highly informational, relatively evaluative, and to some extent organizational and modalizing". By knowing which type of bundles are over-used in learner writing, we learn which chunks of language are easily acquired by students and, in turn, provide practical help for teachers to emphasise other important phrases in the classroom.

In the case of argumentative essays, however, such as those contained in MCSAW, it is certainly hard to avoid personal references and subjective attitudes (Paquot et al., 2013), since learners are explicitly prompted to give their personal opinions. Throughout the study, serious problems have been identified with the amount of tautology or repeated instances across many texts. As discussed in the analyses chapters, there is reason to argue that learners' writing may be influenced by the title prompt or other methods of scaffolding in the classroom that increase their use of certain repetition of sentences. Furthermore, redundant use of language may signal learners' restricted vocabulary range in producing other, varied means of expression. This shows the significant role of essay topics or essay prompts in learner writing and the importance of acquiring adequate vocabulary. As a result, teachers should be aware of the effects of essay topics and forms of scaffolding that may indirectly be over-used in learner writing.

One way in which teachers in the Malaysian ELT classroom can address these and other issues is by using corpora more effectively. This can either be done by using corpus-informed dictionaries (e.g. Collins COBUILD Advanced Learner's Dictionary) and/or accessing online collocation dictionaries (e.g. http://oxforddictionary.so8848.com/). In addition, analysing concordance lines is a simple exercise that can be done in the language classroom. Hence, designing classroom materials based on corpora,[70] i.e. using concordance lines in exercise sheets, not only enables us to use attested data with our learners but enhances their creativity and learner autonomy at the same time (Lee & Swales, 2006). Given this, language teachers may also subsequently assess their students' knowledge (and feedback) about corpus techniques in language tests or assignments. This is particularly useful to cultivate learners'

[70] See examples in Sinclair (2004a).

skills in investigating language themselves, as well as developing their interest in corpus linguistics and other language areas, as Bednarek (2007) has shown.

7.6 Conclusion

Overall, this thesis conducted a corpus-driven contrastive analysis of Malaysian learner argumentative writing (MCSAW) against the reference language variety, LOCNESS. Results indicate that, contrary to their use in LOCNESS, there are particular features that are mostly characteristic of MCSAW, namely the highly over-used modal verb *can*, personal plural pronoun *we*, and referential-type lexical bundles. Furthermore, salient bundles were mostly used in reference to the topic essays (Facebook/living in a hostel); and confirmed the overuse of modal verb *can* and personal plural pronoun *we* described in Chapters 4 and 5, respectively. In addition, *we* + *can* phrases are significantly over-used in MCSAW, and their occurrences are primarily suggestive of the argumentative type essay. More importantly, it was argued that the writing genre and essay topics have greatly influenced the prevalence of certain items in Malaysian learners' writing. Similarly, repeated instances and traces of L1 transfer were also found in MCSAW with the prevalent use of modal verb *can* being mostly attributed to the L1 equivalent *boleh*. In contrast to LOCNESS, MCSAW also showed fewer types of discourse-organising bundles, indicating learners' less varied set of lexico-grammatical repertoire. Despite these differences, there were other areas that showed similar usage between MCSAW and LOCNESS writers. These include using the personal pronoun *we* to achieve solidarity with readers, and bundles that were used similarly (e.g. *is one of the, in my opinion*).

To conclude, the present thesis showed that CIA is indeed an effective way to investigate and compare learner writing. Bearing the above limitations in mind, this thesis advocates for the use of corpora in language research, particularly via the contrastive corpus-driven approach. Apart from reaffirming past findings, results of this thesis goes beyond the descriptive analysis of learner writing and argue that certain features of learner language are still mostly influenced by learners' L1, essay topic and genre of writing. Through its combination of quantitative and qualitative corpus analysis, its new categorisation schemes, and its innovative use of the range and dispersion plot function, this thesis made a significant contribution to the existing body of research on LCR, and provided a range of new insights into Malaysian learner argumentative writing.

References

Aarts, J. (2000). Towards a new generation of corpus-based English grammars. In B. Lewandowska-Tomaszczyk & P. J. Melia (Eds.), *PALC'99*: *Practical applications in language corpora* (pp. 17-36). Frankfurt/Main: Peter Lang.

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

Adolphs, S. & Lin, P. M. S. (2011). Corpus linguistics. In J. Simpson (Ed.), *The Routledge handbook of applied linguistics* (pp. 597–610). Abingdon: Routledge.

Aijmer, K. (1996). *Conversational routines in English: Convention and creativity*. London: Addison Wesley Longman.

Aijmer, K. (2002). Modality in advanced Swedish learners' written interlanguage. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 55-76). Amsterdam: Benjamins.

Akbari, O. (2009). *A corpus based study on Malaysian ESL learners' use of phrasal verbs in narrative compositions*. (Doctoral dissertation). Retrieved from Universiti Putra Malaysia Institutional Repository (UMPIR). (ID No. 11067).

Altenberg, B. & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics*, *22*(2), 173–194.

Ang, L. H., Hajar, A. R., Tan, K. H. & Khazriyati, S. (2011). Collocations in Malaysian English learners' writing: A corpus-based error analysis. *3L: The Southeast Asian Journal of English Language Studies*, *17*(special issue), 31–44.

Anthony, L. (2012). AntConc (Version 3.3.5) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/.

Arshad, A. S., Fauziah H., Mukundan, J., Ghazali K., Sharifah Zainab S. A. R., Juridah M. R. & Vethamani, M. E. (2002). *The English of Malaysian School Students (EMAS) corpus*. Serdang, Malaysia: Universiti Putra Malaysia Press.

Baker, P. (2004). Querying keywords: Questions of difference, frequency and sense in keywords analysis. *Journal of English Linguistics*, *32*(4), 346–359.

Baker, P. (2006). *Using corpora in discourse analysis*. London/New York: Continuum.

Bao, Z. (2010). Must in Singapore English. *Lingua*, *120*(7), 1727–1737.

Barlow, M. (2005). Computer-based analyses of learner language. In G. Barkhuizen & R. Ellis (Eds.), *Analysing learner language* (pp. 335-357). Oxford: Oxford University Press.

Barlow, M. (2011). Corpus linguistics and theoretical linguistics. *International Journal of Corpus Linguistics*, *16*(1), 3–44.

Bednarek, M. (2006). *Evaluation in media discourse: Analysis of a newspaper corpus*. New York/London: Continuum.

Bednarek, M. (2007). Teaching English literature and linguistics using corpus stylistic methods. In C. Cloran & M. Zappavigna (Eds.), *Bridging Discourses*. ASFLA 2007 Online Proceedings. Retrieved from http://www.asfla.org.au/category/asfla2007.

Bednarek, M. (2008a). "An increasingly familiar tragedy": Evaluative collocation and conflation. *Functions of Language*, *15*(1), 7–34.

Bednarek, M. (2008b). *Emotion talk across corpora*. Houndsmills: Palgrave Macmillan.

Berber-Sardinha, T. (2000). Comparing corpora with WordSmith Tools: How large must the reference corpus be? *Proceedings of The Workshop on Comparing corpora - Volume 9*, 7–13. doi: 10.3115/1117729.1117731.

Bestgen, Y. & Granger, S. (2014). Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of Second Language Writing*, *26*, 28–41.

Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes*, *5*(2), 97–116.

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, *14*(3), 275–311.

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

Biber, D., Conrad, S. & Cortes, V. (2003). Lexial bundles in speech and writing: An initial taxonomy. In A. Wilson, P. Rayson & T. McEnery (Eds.), *Corpus linguistics by the lune* (pp. 71-93). Frankfurt/Main: Peter Lang.

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

Biber, D., Conrad, S. & Leech, G. (2002). *Student grammar of spoken and written English*. Harlow, UK: Pearson Education Limited.

Biber, D., Conrad, S. & Reppen, R. (1994). Corpus-based approaches to issues in applied linguistics. *Applied Linguistics*, *15*(2), 169–189.

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

Bondi, M. (2010). Perspectives on keywords and keyness: An introduction. In M. Bondi & M. Scott (Eds.), *Keyness in texts* (pp. 1–18). Amsterdam: John Benjamins Publishing Company.

Botley, S. P. (2010). A corpus-based comparison of idiom use by Malaysian, British and American students. In *International Conference on Science and Social Research (CSSR 2010),* 139–144.

Botley, S. P. (2014). Argument stucture in learner writing: A corpus-based analysis using argument mapping. *Kajian Malaysia*, *32*(1), 45–77.

Botley, S. P., De Alwis, C., Metom, L. & Izza, I. (2005). *CALES: A corpus-based archive of learner English in Sarawak*. Final Project. Sarawak, Malaysia: Unit for Research, Development and Commercialization, Universiti Teknologi MARA.

Botley, S. & Dillah, D. (2007). Investigating spelling errors in a Malaysian learner corpus. *Malaysian Journal of ELT Research*, *3*, 74–93.

Breeze, R. (2007). How personal is this text ? Researching writer and reader presence in student writing using Wordsmith Tools. *Computer Resources for Language Learning, 1*, 14–21.

Brezina, V. & Meyerhoff, M. (2014). Significant or random? A critical review of sociolinguistic generalisations based on large corpora. *International Journal of Corpus Linguistics*, *19*(1), 1–28.

Brezina, V., McEnery, T. & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, *20*(2).

Butler, C. S. (2004). Corpus studies and functional linguistic theories. *Functions of Language, 11*(2), 147–186.

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

Chang, Y. & Swales, J. (1999). Informal elements in English academic writing: Threats or opportunities for advanced non-native speakers? In C. Candlin & K. Hyland (Eds). *Writing: Texts, processes and practices* (pp. 145 – 167). London: Longman.

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

Chen, Y.-H. & Baker, P. (2014). Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics*, 1–33.

Cheng, W. (2012). *Exploring corpus linguistics: Language in action*. London/New York: Routledge.

Cheng, W., Greaves, C., Sinclair, J. M. H. & Warren, M. (2008). Uncovering the extent of the phraseological tendency: Towards a systematic analysis of concgrams. *Applied Linguistics*, *30*(2), 236–252.

Coates, J. (1983). *The semantics of the modal auxiliaries*. Beckenham, Kent: Croom Helm.

Coates, J. (1987). Epistemic modality and spoken discourse. *Transactions of the Philological Society*, *85*(1), 110–131.

Coates, J. (1990). Modal meaning: The semantic–pragmatic interface. *Journal of Semantics,* *7*(1), 53–63.

Cobb, T. (2003). Analyzing late interlanguage with learner corpora: Québec replications of three European studies. *The Canadian Modern Language Review/ La Revue Canadienne Des Langues Vivantes*, *59*(3), 393–424.

Collins, P. (1991). The modals of obligation and necessity in Australian English. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics* (pp. 145-165). New York: Longman.

Cook, G. (2003). *Applied linguistics*. Oxford: Oxford University Press.

Cook, V. (2014). Going beyond the native speaker in language teaching. *TESOL Quarterly*, *33*(2), 185–209.

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

Cowie, A. P. (Ed.). (1998). *Phraseology: Theory, analysis, and applications*. Oxford: Oxford University Press.

Coxhead, A. (2012). Academic vocabulary, writing and English for academic purposes: Perspectives from second language learners. *RELC Journal*, *43*(1), 137–145.

Crompton, P. (2005). "Where", "In Which", and "In That": A corpus-based approach to error analysis. *RELC Journal*, *36*(2), 157–176.

Crystal, D. (1991) *The Cambridge encyclopaedia of language*. Cambridge: Cambridge University Press.

Culpeper, J. (2009). Keyness : Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's Romeo and Juliet. *International Journal of Corpus Linguistics*, *14*(1), 29–59.

Dam-Jensen, H. & Zethsen, K. K. (2008). Translator awareness of semantic prosodies. *Target*, *20*(2), 203–221.

Darina, L. L., Juliana, A. & Norin Norain, Z. A. (2013). A corpus based study on the use of preposition of time "on" and "at" in argumentative essays of form 4 and form 5 Malaysian students. *English Language Teaching*, *6*(9), 128–135.

David, M. K. (2004). Language policies in a multilingual nation: focus on Malaysia. In M. K. David (Ed.), *Teaching of English in second and foreign language settings: focus on Malaysia* (pp. 1-15). Frankfurt: Peter Lang.

De Clerck, B., Delorge, M. & Simon-Vandenbergen, A.-M. (2011). Semantic and pragmatic motivations for constructional preferences: A corpus-based study of provide, supply, and present. *Journal of English Linguistics*, *39*(4), 359–391.

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

De Cock, S. (2011). Preferred patterns of use of positive and negative evaluative adjectives in native and learner speech: An ELT perspective. In A. Frankenberg-Garcia, L. Flowerdew & G. Aston (Eds.), *New trends in corpora and language learning* (pp. 198– 212). London: Continuum.

De Cock, S. & Granger, S. (2004). High frequency words: The bête noire of lexicographers and learners alike. A close look at the verb "make" in five monolingual learners dictionaries of English. In G. Williams & S. Vessier (Eds.), *Proceedings of the 11th Euralex International Congress* (pp. 233–243). Lorient: Université de Bretagne-Sud.

Derewianka, B. (1990). *Exploring how texts work*. Rozelle, New South Wales: Primary English Teaching.

Ebeling, S. O. (2011). Recurrent word-combinations in English student essays. *Nordic Journal of English Studies*, *10*(*1*), 49–76.

Ebeling, S. O. & Hasselgård, H. (2015). Learners' and native speakers' use of recurrent word-combinations across disciplines. In *Learner Corpus Research: LCR2013 Conference Proceedings* (pp. 87–106). doi: 10.15845/bells.v6i0.810.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, R. & Barkhuizen, G. P. (2005). *Analysing learner language*. Oxford: Oxford University Press.

Fan, P. (2010). Lexical acquisition viewed from a contrastive analysis of collocational behavior of near synonyms. *Chinese Journal of Applied Linguistics*, *33*(5), 155–156.

Fawcett, C. L. (2013). *A Corpus-assisted study of Chinese EFL learners' development of academic literacy*. (Doctoral dissertation).

Ferris, D. R. (1994). Rhetorical strategies in student persuasive writing: Differences between native and non-native English speakers. *Research in the Teaching of English*, *28*(1), 45–65.

Flowerdew, J. (2009). Corpora in language teaching. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 327–350). West Sussex: Blackwell Publishing Ltd.

Flowerdew, L. (1998). Integrating 'expert' and 'interlanguage' computer corpora findings on causality: Discoveries for teachers and students. *English for Specific Purposes*, *17*(4), 329–345.

Flowerdew, L. (2001). The exploitation of small learner corpora in EAP materials design. In M. Ghadessy, A. Henry & R. L. Roseberry (Eds.), *Small corpus studies and ELT: Theory and practice* (pp. 363-380). Amsterdam: John Benjamins.

Flowerdew, L. (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: Countering criticisms against corpus-based methodologies. *English for Specific Purposes*, *24*(3), 321–332.

Foo, B. & Richards, C. (2004). English in Malaysia. *RELC*, *35*(2), 229–240.

Gablasova, D. & Brezina, V. (2015). Does speaker role affect the choice of epistemic adverbials in L2 speech? Evidence from the Trinity Lancaster Corpus. In J. Romero-Trillo (Ed.), *Yearbook of corpus linguistics and pragmatics 2015* (pp. 117–136). Springer International Publishing Switzerland.

Gabrielatos, C. & Baker, P. (2006). Representation of refugees and asylum seekers in UK newspapers: Towards a corpus-based analysis. In *Joint Annual Meeting of the British Association for Applied Linguistics and the Irish Association for Applied Linguistics (BAAL/IRAAL 2006): From Applied Linguistics to Linguistics Applied: Issues, Practices, Trends*. Available: http://eprints.lancs.ac.uk/265/.

Gabrielatos, C. & McEnery, T. (2005). Epistemic modality in MA dissertations. In P. A. Fuertes Olivera (Ed.), *Lengua y Sociedad: Investigaciones recientes en lingüística aplicada* (pp. 311–331). Valladolid: Universidad de Valladolid.

Gilquin, G. (2001). The integrated contrastive model: Spicing up your data. *Languages in Contrast: International Journal for Contrastive Linguistics*, *3*(1), 95–123.

Gilquin, G. (2015). Contrastive collostructional analysis: Causative constructions in English and French. *ZAA*, *63*(3), 253–272.

Gilquin, G. & Granger, S. (2015). Learner language. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 418–435). Cambridge: Cambridge University Press.

Gilquin, G. & Paquot, M. (2007). Spoken features in learner academic writing: identification, explanation and solution. In M. Davies, P. Rayson, S. Hunston & P. Danielsson (Eds.), *Proceedings of the Fourth Corpus Linguistics Conference CL2007* (pp. 1-12). UK: University of Birmingham.

Gilquin, G. & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction*, *1*(1), 41–61.

Gilquin, G., Granger, S. & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, *6*(4), 319–335.

Götz, S. & Schilk, M. (2011). Formulaic sequences in spoken ENL, ESL and EFL: Focus on British English, Indian English and learner English of advanced German learners. In J. Mukherjee & M. Hundt (eds.) *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap* (pp. 79–100). Amsterdam: John Benjamins.

Granger, S. (1992). A bird's eye view of learner corpus research. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3-33). Amsterdam/Philadelphia: Benjamins.

Granger, S. (1996). From CA to CIA and back: An integrated contrastive approach to bilingual and learner computerised corpora. In K. Aijmer (Ed.), *Languages in contrast: Text-based cross-linguistic studies* (pp. 37–51). Lund: Lund University Press.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and lexical phrases. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 145–160). Oxford: Clarendon Press.

Granger, S. (2003). Error-tagged learner corpora and CALL: A promising synergy. *CALICO Journal*, *20*(3), 465–480.

Granger, S. (2008). Learner corpora in foreign language education. In N. Van Deusen-Scholl & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (Vol. 4) (pp. 337-351). Springer.

Granger, S. (2012). How to use foreign and second language learner corpora. In A. Mackey & S. G. Gass (Eds), *A guide to research methods in second language acquisition* (pp. 7-29). Malden: Basil Blackwell.

Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research*, *1*(1), 7–24.

Granger, S. & Dumont, A. (2012). *Learner corpora around the world*. Retrieved from http://www.uclouvain.be/en-cecl-lcworld.html

Granger, S. & Paquot, M. (2009). Lexical verbs in academic discourse: A corpus-driven study of expert and learner use. In M. Charles, D. Pecorari & S. Hunston (Eds.), *Academic writing: At the interface of corpus and discourse* (pp. 193–214). London/New York: Continuum.

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

Granger, S. & Rayson, P. (1998). Automatic profiling of learner texts. In S. Granger (Ed.), *Learner English on computer* (pp. 119–131). London/New York: Longman.

Granger, S. & Tribble, C. (1998). Learner corpus data in the foreign language classroom: Form-focused instruction and data- driven learning. In S. Granger (Ed.), *Learner English on computer* (pp. 199–209). London: Longman.

Granger, S. & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes, 15*(1), 17–27.

Granger, S., Gilquin, G. & Meunier, F. (Eds.). (2013). *Twenty Years of Learner Corpus Research: Looking back, Moving ahead. Corpora and Language in Use – Proceedings 1*, Louvain-la-Neuve, Presses universitaires de Louvain.

Gries, S. T. (2006). Exploring variability within and between corpora: Some methodological considerations. *Corpora*, *1*(2), 109–151.

Gries, S. T. (2010). Corpus linguistics and theoretical linguistics: A love-hate relationship? Not necessarily …. *International Journal of Corpus Linguistics, 15*(3), 327–343.

Gries, S. T. (2013). 50-something years of work on collocations: What is or should be next. *International Journal of Corpus Linguistics*, *18*(1), 137–165.

Gries, S. T. & Deshors, S. C. (2014). Using regressions to explore deviations between corpus data and a standard/target: two suggestions. *Corpora*, *9*(1), 109–136.

Gries, S. T. & Divjak, D. (2009). Corpus-based cognitive semantics a contrastive study of phrasal verbs in English and Russian. In V. Evans & S. S. Pourcel (Eds.), *New directions in cognitive linguistics* (pp. 57–75). Amsterdam/Philadelphia: John Benjamins.

Hajar, A. R. (2014). Corpora in language research in Malaysia. *Kajian Malaysia, 32*, 1–16.

Hajar, A. R. (2006). The evolution of Malaysian English: Influences from within. In *The 16th Biennial Conference of the ASAA* (pp. 1-21).

Halliday, M. A. K. & Hasan, R. (1976). *Cohesion in English*. London/New York: Routledge.

Harwood, N. (2005a). "Nowhere has anyone attempted ... In this article I aim to do just that" A corpus-based study of self-promotional I and we in academic writing across four disciplines. *Journal of Pragmatics*, *37*, 1207–1231.

Harwood, N. (2005b). "We do not seem to have a theory ... The theory I present here attempts to fill this gap": Inclusive and exclusive pronouns in academic writing. *Applied Linguistics*, *26*(3), 343–375.

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: a study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied*

Linguistics, *4*(2), 237–260.

Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features*. London/New York: Routledge.

Hinkel, E. (2009). The effects of essay topics on modal verb uses in L1 and L2 academic writing. *Journal of Pragmatics*, *41*(4), 667–683.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics, 19*(1), 24–44.

Hu, C. & Li, X. (2015). Epistemic modality in the argumentative essays of Chinese EFL learners. *English Language Teaching*, *8*(6), 20–31.

Huat, C. M. (2003). Contextualizing language learning: The role of a topic- and genre-specific pedagogic corpus. *TESL Reporter, 36*(2), 42–54.

Hunston, S. (2002a). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Hunston, S. (2002b). Methods in corpus linguistics: Beyond the concordance line. In S. Hunston (Ed.), *Corpora in applied linguistics* (pp. 67–95). Cambridge: Cambridge University Press.

Hunston, S. (2007). Semantic prosody revisited. *International Journal of Corpus Linguistics, 2*(2), 249–268.

Hyland, K. (1990). A genre description of the argumentative essay. *RELC Journal, 21*(1), 66–78.

Hyland, K. (2001). Bringing in the reader: Addressee features in academic articles. *Written Communication*, *18*(4), 549–574.

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

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

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. London/New York: Bloomsbury Publishing.

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

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

Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, *32*, 150–169.

Hyland, K. & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing*, *6*(2), 183–205.

Imm, T. S. (2009). Lexical borrowing from Chinese languages in Malaysian English. *World Englishes*, *28*(4), 451–484.

Imran, Ho, Abdullah. (1993). *The semantics of the modal auxiliaries of Malay*. Kuala Lumpur, Malaysia: Dewan Bahasa Pustaka.

Ishikawa, S. (2007). A Corpus-based study on the vocabulary of English speech presentations by Japanese learners of English and English native speakers. In *Proceedings of the 6th CULI Conference*.

Ishikawa, S. (2011). A new horizon in learner corpus studies: The aim of the ICNALE project. In G. Weir, S. Ishikawa & K. Poonpon. (Eds.), *Corpora and language technologies in teaching, learning and research* (pp. 3–11). Glasgow: University of Strathclyde Press.

Ishikawa, S. (Ed.). (2014). Learner corpus studies in Asia and the world. In *LCSAW2014*. Kobe, Japan: School of Languages & Communication, Kobe University.

Johns, A. M. (1997). *Text, role, and context: Developing academic literacies*. New York: Cambridge University Press.

Kamariah, Y. & Su'ad, A. (2011). Collocational competence among Malaysian undergraduate law students. *Malaysian Journal of ELT Research*, *7*(1), 151-202.

Kennedy, G. (2002). Variation in the distribution of modal verbs in the British National Corpus. In R. Reppen, S. M. Fitzmaurice & D. Biber (Eds.), *Using corpora to explore linguistic variation* (pp. 73–90). Amsterdam: John Benjamins Publishing.

Khojasteh, L. & Kafipour, R. (2012). Have the modal verb phrase structures been well presented in Malaysian English language textbooks? *English Language and Literature Studies*, *2*(1), 35–41.

Knowles, G. & Zuraidah, M. D. (2004). Introducing MACLE: The Malaysian Corpus of Learner English. In *The first national symposium on corpus linguistics: Selected papers*. Wang Longyin & He Anping, Guong Zhou: North East Normal University Press.

Krug, M. G. (2000). *Emerging English modals: A corpus-based study of grammaticalization*. New York: Walter de Gruyter.

Kuo, C.-H. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes*, *18*(2), 121–138.

Laporte, S. (2012). Mind the gap! Bridge between world Englishes and learner Englishes in the making. *English Text Construction*, *5*(2), 264–291.

Laufer, B. & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, *61*(2), 647–672.

Lee, D. Y. W. (2008). Corpora and discourse analysis: new ways of doing old things. In V. K. Bhatia, J. Flowerdew & R. Jones (Eds.), *Advances in discourse studies* (pp. 86–99). London: Routledge.

Lee, D. Y. W. (2010). What corpora are available? In A. O'keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics*. New York: Routledge.

Lee, D. Y. W. & Chen, S. X. (2009). Making a bigger deal of the smaller words: Function words and other key items in research writing by Chinese learners. *Journal of Second Language Writing*, *18*(3), 149–165.

Lee, D. Y. W. & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, *25*, 56–75.

Leech, G. & Coates, J. (1980), Semantic indeterminacy and the modals. In S. Greenbaum et al. (Eds.), *Studies in English Linguistics* (pp. 79-90). London, Longman.

Leech, G. (1992). Corpora and theories of linguistic performance. In J. Svartvik (Ed.), *Directions in Corpus Linguistics: Proceedings of the Nobel Symposium 82* (pp. 105–122). Berlin: Mouton de Gruyter.

Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh: Edinburgh University Press.

Liu, D. (2008). Linking adverbials: An across-register corpus study and its implications. *International Journal of Corpus Linguistics*, *13*(4), 491–518.

Liu, M. & Braine, G. (2005). Cohesive features in argumentative writing produced by Chinese undergraduates. *System*, *33*(4), 623–636.

Lock, G. & Lockhart, C. (1998). Genre in an academic writing class. *Hong Kong Journal of Applied Linguistics*, 47-64.

Lorenz, G. (1999). Learning to cohere: Causal links in native vs. non-native argumentative writing. In W. Bublitz, U. Lenk & E. Ventola (Eds.), *Coherence in spoken and written discourse: How to create it and how to describe it* (pp. 55–75). Amsterdam: John Benjamins.

Luzón, M. J. (2009). The use of we in a learner corpus of reports written by EFL Engineering students. *Journal of English for Academic Purposes*, *8*(3), 192–206.

Mahlberg, M. (2005). *English general nouns: A corpus theoretical approach*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Marchi, A. (2013). *The Guardian on journalism. A corpus-assisted discourse study of self-reflexivity*. (Doctoral dissertation). Lancaster University.

Martin, J. R. (1992). *English text: System and structure*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Martin, J. R. & Rose, D. (2003). *Working with discourse: Meaning beyond the clause*. London: Continuum.

Martinez-Garcia, M. T. & Wulff, S. (2012). Not wrong, yet not quite right: Spanish ESL students' use of gerundial and infinitival complementation. *International Journal of*

Applied Linguistics, *22*(2), 225–244.

McCrostie, J. (2008). Writer visibility in EFL learner academic writing: A corpus-based study. *ICAME Journal*, *32*, 97–114.

McEnery, T. & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.

McEnery, T. & Kifle, N. A (2002). Epistemic modality in argumentative essays of second-language writers. In J. Flowerdew (Ed.). *Academic Discourse*. Harlow: Longman.

McEnery, T. & Wilson, A. (2001). *Corpus linguistics*, (2nd ed.). Edinburgh: Edinburgh University Press.

Mia Emily, A. R., Emma Marini, A. R. & Han Ning, C. (2013). Distribution of articles in written composition among Malaysian ESL learners. *English Language Teaching*, *6*(10), 149–157.

Miller, R. T., Mitchell, T. D. & Pessoa, S. (2016). Impact of source texts and prompts on students' genre uptake. *Journal of Second Language Writing*, *31*, 11–24.

Milton, J. & Tsang, E. (1993). A corpus-based study of logical connectors in EFL students' writing: directions for future research. *Lexis in Studies*, 215–246.

Mindt, D. (1995). *An empirical grammar of the English verb: Modal verbs*. Berlin: Cornelsen.

Mohamed Ismail, A. K., Begi, N. & Vaseghi, R. (2013). A corpus-based study of Malaysian ESL learners' use of modals in argumentative compositions. *English Language Teaching*, *6*(9), 146–157.

Morgan, M. (2011). *A corpus based investigation into the relationship between propositional content and metadiscourse in student essay writing*. (Master's thesis). University of Nottingham.

Mukundan, J. & Kalajahi, S. A. R. (2013). *Malaysian Corpus of Students' Argumentative Writing (MCSAW)*. Victoria, Australia: Lulu Press Inc.

Mukundan, J. & Khojasteh, L. (2011). Modal auxiliary verbs in prescribed Malaysian English textbooks. *Language Teaching*, *4*(1), 79–89.

Mukundan, J. & Menon, S. (2007). Lexical similarities and differences in the mathematics, science and English language textbooks. *K@ta*, *9*(2), 91–111.

Mukundan, J., Khairil Anuar, S., Razalina, I. & Nur Hairunnisa, J. Z. (2013). Malaysian ESL students' syntactic accuracy in the usage of English modal verbs in argumentative writing. *English Language Teaching*, *6*(12), 98–105.

Mukundan, J., Leong Chiew Har, A. & Nimehchisalem, V. (2012). Distribution of articles in Malaysian secondary school English language textbooks. *English Language and Literature Studies*, *2*(2), 62–70.

Neff, J., Ballesteros, F., Dafouz, E., Martínez, F., Rica, J.-P., Díez, M. & Prieto, R. (2004). Formulating writer stance: A contrastive study of EFL learner corpora. In U. Connor & T. A. Upton (Eds.), *Applied corpus linguistics. A multidimensional perspective* (pp. 73–89). Amsterdam/New York: Rodopi.

Neff, J., Dafouz, E., Herrera, H., Martinez, F., Rica, J., Diez, M., Prieto, R. & Sancho, C. (2003). Contrasting learner corpora: the use of modal and reporting verbs in the expression of writer stance. In S. Granger & S. Petch-Tyson (Eds.), *Extending the scope of corpus-based research: New applications, new challenges*. Amsterdam/New York: Rodopi.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English. *Applied Linguistics*, *24*(2), 223–242.

Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. In J. M. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 125–152). Amsterdam: John Benjamins.

Nesselhauf, N. (2009). Co-selection phenomena across New Englishes. *English World-Wide*, *30*(1), 1–26.

Noor Abidah, M. O. & Zaidah, Z. (Eds.). (2008). *Research in English language teaching*. Johor, Malaysia: Penerbit Universiti Teknologi Malaysia.

Noorzan, M. N. (1998). *Word combinations for business English: A study based on a commerce and finance corpus for ESP/ESL applications*. (Doctoral dissertation). Lancaster University.

Nor Hafizah, A., Ong Luyee, E., Indra Gabriel, J. & Kalajahi, S. A. R. (2013). An analysis: The usage of metadiscourse in argumentative writing by Malaysian tertiary level of students. *English Language Teaching*, *6*(9), 83–96.

Normazidah, C. M., Lie, K. Y. & Hazita, A. (2012). Exploring English language learning and teaching in Malaysia. *GEMA Online^{TM} Journal of Language Studies*, *12*(1), 35–51.

Ooi, V. B. Y. (2008). The Lexis of electronic gaming on the web: A sinclairian approach. *International Journal of Lexicography*, *21*(3), 311–323.

Oster, U. (2010). Using corpus methodology for semantic and pragmatic analyses: What can corpora tell us about the linguistic expression of emotions? *Cognitive Linguistics*, *21*(4), 727–763.

Palmer, F. R. (1990). *Modality and the English modals* (2nd ed.). New York: Longman Group UK Limited.

Palmer, F. R. (2001). *Mood and modality*. Cambridge: Cambridge University Press.

Paquot, M. (2010). *Academic vocabulary in learner writing: From extraction to analysis*. London: Continuum.

Paquot, M. (2013). Lexical bundles and L1 transfer effects. *International Journal of Corpus Linguistics*, *18*(3), 391–417.

Paquot, M. & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, *32*, 130–149.

Paquot, M., Hasselgård, H. & Ebeling, S. O. (2013). Writer/reader visibility in learner writing across genres. A comparison of the French and Norwegian components of the ICLE and VESPA learner corpora. In S. Granger, G. Gilquin & F. Meunier (Eds.), *Twenty Years of Learner Corpus Research: Looking back, Moving ahead. Corpora and Language in Use - Proceedings 1* (pp. 377–387). Louvain-la-Neuve: Presses universitaires de Louvain.

Partington, A. (2004). “Utterly content in each other’s company”: Semantic prosody and semantic preference. *International Journal of Corpus Linguistics*, *9*(1), 131–156.

Partington, A. (2011). Phrasal irony: Its form, function and exploitation. *Journal of*

Pragmatics, *43*(6), 1786–1800.

Peromingo, J. R. (2012). Corpus analysis and phraseology: Transfer of multi-word units. *Linguistics and the Human Sciences, 6*(1-3), 321–343.

Petch-Tyson, S. (1998). Writer/reader visibility in EFL written discourse. In S. Granger (Ed.), *Learner English on computer* (pp. 107–118). London/New York: Longman.

Pravec, N. A. (2002). Survey of learner corpora. *ICAME Journal*, *26*(1), 81–114.

Qian, L. (2010). *A comparative genre analysis of English argumentative essays written by English major and non-English major students in an EFL context*. (Master's thesis). Suranaree University of Technology.

Quirk, R. & Greenbaum, S. (1975). *A concise grammar of contemporary English*. New York: Harcourt Brace Jovanovich.

Rankin, T. (2012). The transfer of V2: inversion and negation in German and Dutch learners of English. *International Journal of Bilingualism*, *16*(1), 139–158.

Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, *13*(4), 519–549.

Ringbom, H. (1998). Vocabulary frequencies in advanced learner English: a cross-linguistic approach. In S. Granger (Ed.), *Learner English on computer* (pp. 41-52). London: Longman.

Römer, U. (2004). A corpus-driven approach to modal auxiliaries and their didactics. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 185–199). Amsterdam: John Benjamins Publishing.

Römer, U. (2006). Pedagogical applications of corpora: Some reflections on the current scope and a wish list for future developments. *Zeitschrift Fur Anglistik Und Amerikanistik*, *54*(2), 121–134.

Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics*, *7*(2009), 140–162.

Saadiyah, D. (2009). The current situation and issues of the teaching of English in Malaysia.

Ritsumeikan Studies in Language and Culture, *22*(*1*), 19-28.

Salazar, D. & Verdaguer, I. (2009). Polysemous verbs and modality in native and non-native argumentative writing: a corpus-based study. *International Journal of English Studies*, 209–219.

Schneer, D. (2014). Rethinking the argumentative essay. *TESOL Journal*, *5*(4), 619–653.

Scott, M. (1997). PC analysis of key words – and key key words. *System*, *25*(2), 233-245.

Scott, M. (1999). *WordSmith Tools Help Manual*. (Version 3.0). Oxford: Oxford University Press.

Scott, M. (2001). Comparing corpora and identifying key words, collocations, and frequency distributions through the WordSmith Tools suite of computer programs. In M. Ghadessy, A. Henry & R. L. Roseberry (Eds.), *Small corpus studies and ELT: Theory and practice* (pp. 47–67). Amsterdam: John Benjamins.

Scott, M. (2008). Developing WordSmith. *International Journal of English Studies*, *8*(1), 95–106.

Scott, M. (2012). WordSmith Tools. (Version 6.0). [Computer Software]. Stroud: Lexical Analysis Software. Available from http://lexically.net/wordsmith/downloads/

Scott, M. (2015). WordSmith Tools Help. Stroud: Lexical Analysis Software.

Scott, M. & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. Amsterdam/Philadelphia: John Benjamins Publishing.

Selinker, L. (2014). Interlanguage 40 years on. Three themes from here. In Z. Han & E. Tarone (Eds.), *Interlanguage. Forty years later* (pp. 221-246). Amsterdam/Philadelphia: John Benjamins.

Simpson, R. (2004). Stylistic features of academic speech: The role of formulaic expressions. In T. Upton & U. Connor (Eds.), *Discourse in the professions: Perspectives from corpus linguistics* (pp. 37–64). Amsterdam: John Benjamins.

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

Sinclair, J. (1996). The search for units of meaning. *Textus*, *9*, 75–106.

Sinclair, J. (2004a). *Trust the text. Language, corpus and discourse*. London/New York: Routledge.

Sinclair, J. (2004b). Intuition and annotation – the discussion continues. In K. Aijmer & B. Altenberg (Eds.), *Advances in corpus linguistics*. *Papers from the 23rd International conference on English Language Research on Computerised Corpora* (ICAME 23) (pp. 39–59). Amsterdam: Rodopi Press.

Sinclair, J. (Ed.), (1987). *Looking up*. Collins, London.

Sinclair, J., Jones, S., Daley, R. & Krishnamurthy, R. (2004). *English collocation studies: The OSTI report*. London/New York: Continuum.

Siti Aeisha, J. & Hajar, A. R. (2014). Corpus research in Malaysia: A bibliographic analysis. *Kajian Malaysia*, *32*, 17–43.

Smith, K. A. & Nordquist, D. (2012). A critical and historical investigation into semantic prosody. *Journal of Historical Pragmatics*, *13*(2), 291–312.

Staples, S. & Reppen, R. (2016). Understanding first-year L2 writing: A lexico-grammatical analysis across L1s, genres, and language ratings. *Journal of Second Language Writing*, *32*, 17–35.

Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language*, *2*(1), 23–55.

Stubbs, M. (2003). Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics*, *7*(2), 215–244.

StudyMalaysia.com. (2015). A glance at the Malaysian education system. Retrieved from https://www.studymalaysia.com/education/higher-education-in-malaysia/a-glance-at-the-malaysian-education-system

Sweetser, E. (1990). *From etymology to pragmatics*. Cambridge: Cambridge University Press.

Tang, R. & John, S. (1999). The “I” in identity: Exploring writer identity in student academic writing through the first person pronoun. *English for Specific Purposes, 18,* S23–S39.

Taylor, C. (2008). What is corpus linguistics? What the data says. *ICAME Journal*, 179–200.

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

Thewissen, J. (2013). Capturing L2 accuracy developmental patterns: Insights from an error-tagged EFL learner corpus. *Modern Language Journal*, *97*(S1), 77–101.

Thompson, G. & Hunston, S. (2006). *System and corpus: Exploring connections*. London: Equinox.

Tognini-Bonelli, E. (2000). Corpus classroom currency. *Darbai Ir Dienos*, *24*, 205–243.

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

Tribble, C. (1991) Some uses of electronic text in English for academic purposes. In J. Milton & K. Tong (Eds.), *Text analysis in computer assisted language learning* (pp. 4-14). Hong Kong: The Hong Kong University of Science and Technology.

Tribble, C. (2011). Revisiting apprentice texts: Using lexical bundles to investigate expert and apprentice performances in academic writing. In F. Meunier, S. De Cock, G. Gilquin & M. Paquot (Eds.), *A taste for corpora: In honour of Sylviane Granger* (pp. 85–108). Amsterdam: John Benjamins Publishing Company.

Veel, R. (1997). Learning how to mean – scientifically speaking: Apprenticeship into scientific discourse in the secondary school. In F. Christie & J. R. Martin (Eds.), *Genre and institution: Social processes in the workplace and school* (pp. 161-195). London: Continuum.

Vethamani, M. E., Umi Kalthom, A. M. & Akbari, O. (2010). Students use of modals in their written work: Compensation strategies and simplification features. *Studies in Languages and Language Teaching*, *14*(2), 13–26.

Walker, C. P. (2011). A corpus-based study of the linguistic features and processes which influence the way collocations are formed: Some implications for the learning of collocations. *TESOL Quarterly*, *45*(2), 291–312.

Warner, A. (1993). *English auxiliaries*. Cambridge: Cambridge University Press.

Webb, S. (2010). A corpus driven study of the potential for vocabulary learning through watching movies. *International Journal of Corpus Linguistics*, *15*(4), 497–519.

Wingate, U. (2012). “Argument!” helping students understand what essay writing is about. *Journal of English for Academic Purposes*, *11*, 145–154.

Wray, A. (1999). Formulaic language in learners and native speakers. *Language Teaching*, *32*, 213–231.

Xiao, R. Z., McEnery, T. & Qian, Y. (2006). Passive constructions in English and Chinese: A corpus-based contrastive study. *Languages in Contrast*, *1*, 109–149.

Xiao, R. & McEnery, T. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics*, *27*(1), 103–129.

Xu, J. & Liang, M. (2012). A tale of two C’s : Comparing English varieties with Crown and CLOB (the 2009 Brown family corpora). *ICAME Journal*, *37*, 175–184.

Yang, W. & Sun, Y. (2012). The use of cohesive devices in argumentative writing by Chinese EFL learners at different proficiency levels. *Linguistics and Education*, *23*(1), 31–48.

Yunisrina, Q. Y. (2009). A corpus-based linguistics analysis on written corpus: Colligation of “TO” and “FOR.” *Journal of Language and Linguistic Studies*, *5*(2).

Zarifi, A. & Mukundan, J. (2012). Phrasal verbs in Malaysian ESL textbooks. *English Language Teaching*, *5*(5), 9–18.

Zhang, W. (2009). Semantic prosody and ESL / EFL vocabulary pedagogy. *TESL Canada Journal/Revue TESL Du Canada*, *26*(2), 1–12.

Zhu, W. (2001). Performing argumentative writing in English: Difficulties processes and strategies. *TESL Canada Journal/Revue TESL Du Canada*, *19*(1), 34–50.

Zuraidah, M. D., Knowles, G. & Fatt, C. K. (2010). Nationhood and Malaysian identity: A corpus-based approach. *Text and Talk*, *30*(3), 267–287.

Appendix

Table A4.1: Keywords list[71]

N	Key word	Freq.	Texts	RC. Freq.	RC. %	Keyness
1	facebook	7487	477	0		12517.86
2	can	4124	495	1116	0.34	3773.52
3	we	3148	429	925	0.29	2727.58
4	friends	1412	414	50	0.02	2155.75
5	advantages	1149	441	28		1796.04
6	disadvantages	1034	409	14		1654.76
7	our	1739	370	584	0.18	1364.95
8	us	1183	347	212	0.07	1315.06
9	information	891	343	104	0.03	1135.40
10	using	869	318	95	0.03	1125.44
11	use	1271	401	362	0.11	1117.51
12	social	982	380	174	0.05	1095.66
13	your	886	218	122	0.04	1078.05
14	share	606	293	30		894.06
15	people	2310	454	1569	0.48	788.66
16	their	2254	459	1540	0.48	761.22
17	also	1577	457	861	0.27	754.34
18	know	757	349	194	0.06	707.87
19	networking	429	237	0		705.04
20	you	1180	230	543	0.17	693.06
21	time	1125	375	497	0.15	689.88
22	connect	417	248	2		677.99
23	users	435	205	17		655.77
24	get	809	333	276	0.09	624.99
25	friend	427	199	29		604.01
26	network	391	188	20		573.66
27	students	677	233	237	0.07	512.07
28	online	310	176	0		508.62
29	nowadays	322	256	20		460.66
30	profile	279	116	0		457.46
31	communicate	310	175	17		450.56
32	with	2214	474	1909	0.59	448.51
33	account	313	170	20		445.95
34	internet	284	179	9		434.18
35	business	455	178	114	0.04	429.91
36	medium	274	175	8		421.11
37	hostel	250	29	0		409.62
38	besides	267	198	19		373.98
39	communication	278	177	25		373.71
40	user	228	118	2		366.13
41	news	308	172	45	0.01	366.05
42	chat	211	144	0		345.28
43	student	307	121	56	0.02	336.72
44	group	346	171	87	0.03	325.76

[71] Negative keywords are highlighted.

45	stay	312	126	69	0.02	314.05
46	world	717	339	426	0.13	302.81
47	fake	192	85	6		292.95
48	easily	265	196	53	0.02	279.25
49	advantage	236	133	36	0.01	275.99
50	conclusion	244	238	41	0.01	275.81
51	page	178	115	5		273.30
52	make	693	302	433	0.13	270.61
53	lot	307	184	90	0.03	263.01
54	other	939	379	711	0.22	257.42
55	popular	248	203	52	0.02	255.65
56	site	161	100	3		252.17
57	or	1423	425	1286	0.40	252.07
58	many	1120	381	925	0.29	251.94
59	find	388	215	166	0.05	244.44
60	chatting	152	108	1		244.37
61	about	802	341	578	0.18	243.07
62	website	148	96	0		241.37
63	post	203	128	33	0.01	231.89
64	addicted	150	109	5		227.25
65	waste	195	135	36	0.01	212.06
66	others	442	233	243	0.07	207.93
67	around	335	220	147	0.05	205.47
68	study	223	144	58	0.02	205.25
69	games	204	122	45	0.01	204.90
70	video	138	102	6		204.23
71	old	332	217	152	0.05	194.67
72	updates	119	79	0		193.55
73	disadvantage	131	90	6		192.75
74	contact	173	127	30		192.63
75	give	341	172	164	0.05	189.28
76	moreover	130	109	7		187.88
77	daily	160	134	24		187.46
78	will	1188	344	1116	0.34	186.99
79	than	641	332	480	0.15	179.45
80	status	158	100	26		179.22
81	it	2751	481	3219	0.99	177.51
82	help	314	184	150	0.05	175.53
83	latest	120	87	6		174.72
84	pages	107	66	1		170.19
85	gives	188	128	51	0.02	168.46
86	because	946	377	854	0.26	167.59
87	teenagers	149	90	27		162.78
88	furthermore	135	127	19		161.10
89	income	166	52	40	0.01	158.97
90	them	700	304	581	0.18	155.05
91	relationship	168	110	44	0.01	153.54
92	good	461	263	316	0.10	153.13
93	fan	92	63	1		145.47
94	product	149	93	35	0.01	144.45
95	beside	93	73	2		143.65
96	activities	136	94	27		142.80
97	more	1156	426	1169	0.36	141.63
98	privacy	112	70	13		140.77

99	sharing	100	80	7		138.90
100	benefit	166	116	50	0.02	138.84
101	easy	182	132	63	0.02	137.59
102	spread	119	88	19		135.94
103	twitter	84	69	0		135.84
104	addicting	84	54	1		132.29
105	update	84	67	1		132.29
106	family	334	202	208	0.06	130.45
107	famous	96	83	8		129.38
108	agree	168	119	57	0.02	128.89
109	meet	118	94	23		124.64
110	promote	118	87	23		124.64
111	free	237	168	119	0.04	124.61
112	so	774	342	723	0.22	123.59
113	anywhere	112	103	20		122.57
114	opinion	178	145	69	0.02	122.28
115	photos	79	52	2		120.61
116	wisely	79	66	2		120.61
117	from	1130	399	1187	0.37	119.56
118	pictures	91	70	9		118.37
119	it's	216	115	105	0.03	117.92
120	opportunity	147	112	47	0.01	117.71
121	addition	111	103	22		116.27
122	sites	82	45	5		115.77
123	create	173	144	70	0.02	114.21
124	touch	90	70	10		113.93
125	like	444	253	346	0.11	113.71
126	personal	200	122	94	0.03	113.38
127	wasting	78	54	4		112.39
128	valuable	99	63	16		112.25
129	easier	135	105	41	0.01	111.96
130	most	537	302	457	0.14	111.39
131	advertise	70	43	1		109.23
132	connecting	68	55	1		105.94
133	become	388	236	294	0.09	105.53
134	priority	88	32	12		105.29
135	connected	79	59	7		104.83
136	spend	131	101	42	0.01	104.51
137	feelings	130	67	42	0.01	103.12
138	among	155	117	63	0.02	101.74
139	custom	67	66	2		100.89
140	private	102	84	23		100.39
141	best	274	169	179	0.06	98.66
142	face	158	74	68	0.02	98.08
143	marketing	73	50	6		98.08
144	new	387	218	303	0.09	97.95
145	bad	246	165	153	0.05	95.98
146	firstly	108	101	30		94.62
147	low	130	44	48	0.01	92.77
148	expenses	64	29	3		92.69
149	products	127	86	46	0.01	92.07
150	precious	67	56	5		91.34
151	upload	57	49	0		91.33
152	through	371	210	295	0.09	90.61

153	zuckerberg	56	49	0		89.68
154	especially	198	150	112	0.03	88.87
155	cons	65	55	5		88.10
156	mark	70	63	9		84.77
157	secondly	84	79	18		84.43
158	feedback	54	52	1		82.89
159	technology	154	96	76	0.02	82.22
160	ways	169	108	90	0.03	82.16
161	save	112	90	40	0.01	81.97
162	example	366	220	302	0.09	81.81
163	harm	88	51	22		81.74
164	want	289	174	217	0.07	79.90
165	sometimes	122	95	50	0.02	79.14
166	updated	51	44	1		77.96
167	knowledge	163	99	88	0.03	77.78
168	popularity	81	67	19		77.74
169	stalk	48	45	0		76.50
170	message	77	55	17		76.12
171	homework	61	53	7		75.89
172	brings	101	80	35	0.01	75.59
173	studying	69	60	12		75.52
174	finding	85	66	23		75.32
175	wall	82	58	21		75.01
176	id	47	26	0		74.85
177	addiction	65	53	10		74.27
178	studies	101	63	36	0.01	73.88
179	photo	46	35	0		73.20
180	harming	46	46	0		73.20
181	any	430	240	395	0.12	71.90
182	browsing	45	40	0		71.55
183	don't	209	125	141	0.04	70.89
184	useful	87	69	27		70.49
185	some	548	240	547	0.17	69.94
186	click	44	41	0		69.91
187	platform	44	29	0		69.91
188	cost	126	98	60	0.02	69.89
189	important	274	207	214	0.07	69.48
190	smart	53	48	5		68.69
191	facebook's	43	37	0		68.26
192	assignment	45	37	1		68.09
193	rent	54	26	6		67.47
194	relatives	50	35	4		66.77
195	manage	49	37	4		65.16
196	things	225	147	165	0.05	65.13
197	benefits	113	85	52	0.02	64.99
198	log	41	35	0		64.96
199	videos	43	38	1		64.80
200	send	62	45	12		64.76
201	phone	60	46	11		64.05
202	need	283	170	234	0.07	62.75
203	helping	68	66	17		62.75
204	properly	54	46	8		62.14
205	abroad	54	46	8		62.14
206	offices	57	30	10		61.83

207	depends	57	54	10		61.83
208	application	57	42	10		61.83
209	add	63	54	14		61.74
210	avoid	74	60	22		61.51
211	they	1632	397	2079	0.64	61.20
212	pros	48	46	5		60.66
213	malaysia	38	33	0		60.02
214	exams	42	34	2		59.89
215	customers	57	38	11		59.45
216	biggest	57	43	11		59.45
217	all	825	368	946	0.29	58.49
218	assignments	46	38	5		57.46
219	lastly	46	43	5		57.46
220	depend	54	39	10		57.21
221	block	41	39	3		55.17
222	insult	35	34	0		55.07
223	anytime	37	34	1		54.94
224	do	640	304	708	0.22	53.88
225	helps	89	61	39	0.01	53.64
226	keep	149	118	97	0.03	53.51
227	without	269	187	232	0.07	53.48
228	carefully	43	40	5		52.67
229	picture	80	63	32		52.61
230	country	208	126	163	0.05	52.15
231	call	77	59	30		51.82
232	advertisement	40	26	4		50.65
233	gather	48	41	9		50.38
234	nutshell	32	32	0		50.13
235	comment	43	39	6		50.01
236	careful	55	45	14		49.90
237	networks	63	42	20		49.78
238	era	44	42	7		49.03
239	via	50	31	11		48.86
240	gain	99	74	52	0.02	48.50
241	laptop	31	26	0		48.48
242	harass	35	35	2		48.47
243	instance	73	61	29		48.19
244	connection	38	34	4		47.45
245	foremost	36	36	3		47.08
246	wastes	33	32	2		45.21
247	minimize	29	28	0		45.19
248	discuss	74	56	32		44.99
249	faster	55	44	17		44.09
250	front	66	50	26		43.77
251	limit	67	36	27		43.52
252	discussion	57	39	19		43.28
253	friendship	37	26	5		43.16
254	tool	58	52	20		42.92
255	strongly	52	47	16		41.73
256	button	36	28	5		41.59
257	college	136	85	97	0.03	41.25
258	just	351	205	358	0.11	41.18
259	culture	118	105	78	0.02	41.07
260	for	2273	472	3145	0.97	40.76

261	convenient	40	35	8		40.50
262	publicly	30	30	2		40.34
263	yahoo	26	26	0		40.25
264	thing	129	103	91	0.03	39.93
265	place	190	149	161	0.05	39.36
266	date	60	50	24		39.10
267	projects	39	37	8		38.99
268	sell	68	59	31		38.99
269	happen	104	74	66	0.02	38.68
270	advertising	40	30	9		38.29
271	billion	40	38	9		38.29
272	media	115	78	78	0.02	38.21
273	too	228	145	209	0.06	38.05
274	tradition	69	68	33	0.01	37.51
275	everyone	154	125	122	0.04	37.46
276	marks	30	26	3		37.44
277	playing	71	60	35	0.01	37.28
278	messages	28	26	2		37.09
279	someone	141	101	108	0.03	36.93
280	applications	26	25	1		36.91
281	my	266	166	260	0.08	36.22
282	drastically	44	44	13		36.06
283	creating	62	58	28		35.80
284	entertainment	65	52	31		35.38
285	busy	32	29	5		35.34
286	colleague	25	25	1		35.28
287	anything	109	85	75	0.02	35.22
288	trouble	46	41	15		35.21
289	groups	101	69	68	0.02	33.89
290	that's	45	44	15		33.84
291	proper	45	41	15		33.84
292	long	184	130	164	0.05	33.20
293	interact	35	32	8		33.01
294	actually	121	83	91	0.03	32.85
295	happening	55	54	25		31.39
296	what's	38	38	11		31.34
297	talk	72	59	41	0.01	31.30
298	fine	39	37	12		30.89
299	features	39	28	12		30.89
300	comments	30	27	6		29.91
301	tend	61	41	32		29.48
302	every	197	144	187	0.06	29.41
303	read	71	57	42	0.01	29.17
304	largest	36	31	11		28.53
305	focus	63	37	35	0.01	28.23
306	improve	65	51	37	0.01	28.17
307	created	92	69	65	0.02	28.14
308	almost	108	105	83	0.03	27.89
309	attract	34	34	10		27.57
310	lost	113	91	90	0.03	26.94
311	newspaper	42	36	17		26.65
312	access	51	39	25		26.62
313	activity	44	32	19		26.24
314	next	97	90	73	0.02	26.13

315	books	34	32	11		25.77
316	disagree	38	26	15		24.55
317	current	52	42	28		24.11
318	doing	117	83	99	0.03	24.05
319	believe	68	58	221	0.07	-24.32
320	show	47	45	173	0.05	-24.49
321	article	27	25	124	0.04	-24.56
322	over	119	90	333	0.10	-24.68
323	role	28	27	127	0.04	-24.76
324	before	69	57	225	0.07	-24.96
325	is	3443	499	6307	1.95	-25.23
326	point	39	29	159	0.05	-26.74
327	but	610	327	1298	0.40	-27.07
328	back	45	39	178	0.05	-28.60
329	those	110	83	326	0.10	-28.63
330	cases	29	25	140	0.04	-29.66
331	job	29	25	141	0.04	-30.14
332	what	282	171	687	0.21	-30.60
333	change	33	32	155	0.05	-31.82
334	single	29	26	150	0.05	-34.50
335	down	28	26	149	0.05	-35.34
336	less	37	30	177	0.05	-37.31
337	order	45	34	198	0.06	-37.49
338	may	167	91	486	0.15	-40.80
339	number	40	33	200	0.06	-44.57
340	years	74	64	293	0.09	-47.64
341	end	38	35	204	0.06	-49.19
342	that	2514	484	4919	1.52	-49.60
343	lives	48	38	233	0.07	-50.28
344	suicide	26	26	175	0.05	-51.68
345	only	274	190	756	0.23	-54.20
346	should	279	161	771	0.24	-55.57
347	not	1109	403	2403	0.74	-57.26
348	her	149	93	495	0.15	-58.05
349	at	392	219	1017	0.31	-58.97
350	feel	59	51	281	0.09	-59.42
351	no	223	151	674	0.21	-63.08
352	had	182	101	590	0.18	-65.60
353	public	46	41	260	0.08	-66.33
354	children	92	57	382	0.12	-67.21
355	did	35	33	233	0.07	-68.56
356	seen	29	25	218	0.07	-70.08
357	out	161	122	569	0.18	-76.06
358	human	57	45	320	0.10	-81.33
359	case	29	26	243	0.07	-83.88
360	two	50	45	309	0.10	-85.81
361	being	151	115	572	0.18	-87.01
362	fact	43	40	306	0.09	-95.09
363	which	386	221	1137	0.35	-99.66
364	however	136	107	591	0.18	-111.65
365	made	38	34	327	0.10	-115.09
366	does	65	51	415	0.13	-118.88
367	him	59	53	408	0.13	-124.50
368	she	65	40	435	0.13	-129.51

369	could	138	76	635	0.20	-129.71
370	been	186	123	793	0.24	-146.11
371	these	202	138	839	0.26	-148.77
372	into	64	59	489	0.15	-160.42
373	society	40	36	423	0.13	-165.98
374	an	316	219	1246	0.38	-204.11
375	a	3034	494	6846	2.11	-212.81
376	be	1078	386	3196	0.99	-287.73
377	were	62	41	735	0.23	-304.02
378	#	303	158	1689	0.52	-431.69
379	was	194	104	1555	0.48	-529.93
380	would	122	76	1461	0.45	-608.95
381	of	4082	496	10729	3.31	-674.43
382	his	85	53	1564	0.48	-746.93
383	he	65	44	2186	0.67	-1168.37
384	the	7526	508	21105	6.51	-1696.82

Table A4.2: Key Keywords list

Key KeyWords List			
N	**KW**	**Texts**	**%**
1	facebook	93	110.71
2	can	59	70.24
3	friends	54	64.29
4	we	43	51.19
5	advantages	42	50
6	disadvantages	37	44.05
7	networking	30	35.71
8	online	27	32.14
9	our	27	32.14
10	connect	23	27.38
11	your	20	23.81
12	share	19	22.62
13	you	19	22.62
14	profile	18	21.43
15	using	17	20.24
16	us	15	17.86
17	students	14	16.67
18	hostel	13	15.48
19	information	13	15.48
20	stay	13	15.48
21	their	13	15.48
22	chat	12	14.29
23	fake	12	14.29
24	income	12	14.29
25	network	12	14.29

Table A4.3: Significance test (Keyness) of modal verbs in MCSAW and LOCNESS texts

Positive/negative	**Modal**	**MCSAW Normalised/ raw**	**LOCNESS**	**Keyness value**
Positive	Will	0.60 (1,188)	0.34 (1,116)	186.99
Negative	Would	0.06 (122)	0.45 (1,461)	608.95
Positive	Can	2.12 (4,178)	0.34 (1,116)	3773.52
Negative	Could	0.07 (139)	0.20 (635)	129.71
Negative	Should	0.14 (279)	0.24 (771)	55.57
	Must	0.08 (153)	0.10 (322)	-
Negative	May	0.08 (167)	0.15 (486)	40.80
	Might	0.05 (95)	0.03 (85)	-
	Shall	0.00 (2)	0.00 (11)	-

Table A6.1: 4-word and 3-word referential bundles

Category	Sub-category		4-word bundles	3-word bundles
Referential bundles	Identification/focus bundles		*is one of the* (104), *one of the biggest* (26), *there are a lot* (24)	*there are some* (47)
	Bundles specifying attributes of following nouns (including quantities)			
		Not incorporating the specified entity	*have a lot of* (30), *are a lot of* (25)	*the popularity of* (51), *most of the* (56), *most of your (34)*
		Incorporating the specified entity	*most of the people* (24), *a lot of time* (24)	-
	Bundles specifying attributes of preceding entities			
		Not incorporating the specified entity	*of the most popular* (26)	-
		Incorporating the specified entity	*people around the world* (107), *people in the world* (31)[72]	-
	Time/place-text-deixis[73]		*all over the world* (46), *anywhere in the world* (61), *all around the world* (37)	*in this world* (44), *in our life* (27), *of all time* (39), *in front of* (40)
	Imprecision bundles		*in many ways and* (50)	*and so on* (50), *and many more* (46), *in other ways* (42), *in many ways* (64)
	Other referential bundles		-	*friends and family* (28), *him or her* (35), *we do not* (43), *who works in* (26), *by creating a* (38), *to stay in* (31), *to communicate with* (99), *we use it* (41), *to use it* (47), *with their friends* (67), *as a student* (28), *face to face* (31), *their time in* (35), *for us to* (54), *to any other* (41), *with each other* (38)

[72] Since these bundles contain a place reference, they could also be identified as place deixis but occur here with a preceding noun.

[73] Some of these could also be identified as bundles specifying attributes of preceding entities (e.g. *in our life*, *of all time*), but were classified as time/place/text-deixis here because they contain reference to time and place.

Table A6.2: 4-word and 3-word discourse organising bundles

Category	Sub-category	4-word bundles	3-word bundles
Discourse organizers	Topic introduction/focus	-	*first of all* (58), *first and foremost* (34)
	Topic elaboration/clarification	-	*is a social* (39), *have their own* (41), *for example if* (28), *then it is* (29)
	Inferential	-	*as a conclusion* (47)*, it is because* (26), *this is because* (73)