A CORPUS-DRIVEN ANALYSIS OF MALE AND FEMALE WEBLOG USERS IN AN ESL CONTEXT

HUANG WEN JIE

FACULTY OF LANGUAGES AND LINGUISTICS
UNIVERSITY OF MALAYA
KUALA LUMPUR

2015

A CORPUS-DRIVEN ANALYSIS OF MALE AND FEMALE WEBLOG USERS IN AN ESL CONTEXT

HUANG WEN JIE

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

FACULTY OF LANGUAGES AND LINGUISTICS
UNIVERSITY OF MALAYA
KUALA LUMPUR

2015

UNIVERSITY OF MALAYA
ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **HUANG WEN JIE** I.C/Passport No: **G46802324**

Registration/ Matric No: **TGB110004**

Name of Degree: **Master of English as a Second Language**

Title of Project Paper/Research Report/**Dissertation**/ Thesis ("this work"):
A CORPUS-DRIVEN ANALYSIS OF MALE AND FEMALE WEBLOG USERS IN AN ESL CONTEXT

Field of Study: **Sociolinguistics**

I do solemnly and sincerely declare that:

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

Candidate's Signature Date

Subscribed and solemnly declared before,

Witness's Signature Date

Name:
Designation:

ABSTRACT

Weblogs ushered a new epoch for Malaysian ESL learners ever since its first entry to Malaysia in 1998 (Hopkins, 2010). Since then, scholars in the field of English as a Second Language (ESL) have paid considerable attention to blogging, especially those who teach writing courses (Mah and Er, 2009; Mah and Liaw, 2011; Gedera, 2012; Kaur, Ganapathy, and Sidhu, 2012; Yunus, Kiing, and Salehi, 2013; Ubaidullah, Mahadi, and Ching, 2013). However, little research has been undertaken from a sociolinguistic perspective, let alone gender-oriented to analyse adults' weblogs. To fill the gap and promote online communication in an ESL community, the present study uses a corpus-driven approach and via a corpus tool, namely, Wmatrix (Rayson, 2013) to tag the key semantic domains (primary) and parts of speech (auxiliary) in 200 ESL weblogs produced by 100 Malaysian females and 100 males from http://www.blogmalaysia.com/. Quantitative results reveal gender differences identified by the log-likelihood value in females' and males' use of key semantic domains and parts of speech. Findings reveal that certain semantic domains and parts of speech are more significant and predominantly found in the female weblogs compared to the males. The findings are also attributed to socio-cultural contexts. Future research can be conducted with other groups of bloggers and a larger corpus is also recommended.

Keywords: Corpus linguistics, semantic domain, part of speech, weblog, gender, Wmatrix

ABSTRAK

Kemunculan weblog di Malaysia pada tahun 1998 (Hopkins, 2010) telah mewujudkan satu zaman yang baru untuk pembelajaran bahasa Inggeris sebagai bahasa kedua (ESL) di Malaysia sejak. Semenjak itu, para sarjana dalam pengajian bidang pembelajaran bahasa Inggeris sebagai bahasa kedua (ESL), khususnya mereka yang mengajar kursus penulisan, telah menumpukan lebih banyak perhatian kepada blog, (Mah dan Er, 2009; Mah dan Liaw, 2011; Gedera, 2012; Kaur, Ganapathy, dan Sidhu, 2012; Yunus, Kiing, and Salehi, 2013; Ubaidullah, Mahadi, dan Ching, 2013). Walau bagaimanapun, kajian yang dijalankan dari segi linguistik sosial masih kurang, apatah lagi penyelidikan yang bertumpu pada pengaruh jantina terhadap weblog oleh golongan dewasa. Untuk mengisi jurang penyelidikan ini serta mendorong komunikasi Internet dalam komuniti ESL, penyelidikan ini telah dijalankan menggunakan pendekatan korpus serta perisian korpus, Wmatrix (Rayson, 2013), untuk mengenalpasti domain semantik utama (primer) dan *parts of speech* (tambahan) dalam 200 weblog ESL yang dihasilkan oleh 100 orang perempuan dan 100 orang lelaki dari http://www.blogmalaysia.com/. Keputusan kuantitatif menunjukkan perbezaan di antara jantina yang dikenal pasti dengan menggunakan nilai log-kemungkinan terhadap pilihan domain semantik utama dan *parts of speech* wanita dan lelaki. Dapatan kajian menunjukkan bahawa terdapat domain semantik utama dan *parts of speech* yang lebih signifikan dan lebih kerap didapati dalam weblog wanita berbanding dengan lelaki. Penemuan kajian juga dikaitkan dengan konteks sosio-budaya. Kajian masa depan boleh dijalankan terhadap kumpulan blogger yang lebih spesifik, dan korpora yang lebih besar juga digalakkan.

Kata kunci: linguistik Korpus, domain semantik, parts of speech, weblog, jantina, Wmatrix

ACKNOWLEDGEMENTS

First and formost, I would like to express my deepest appreciation to all my families who always support me very much, especially my parents, and my fiance. I love you all.

Secondly, I will express my gratitude to my dear supervisor Dr. Sheena Kaur Jaswant Singh. Thanks for her patient instruction, support, guidance, valuable time and wise suggestions for my dissetation.

Thirdly, I would like to express my sincere grateful to all the lectures and teachers who has taught me from the beginning of my education until now. Thanks for teaching me the various sections of specialized knowledge and the way of living a life. Especially, I would like to thank Associate Professor Dr. Jariah Binti Mohd Jan and Dr. Siti Zaidah Binti Zainuddin for their invaluable feedback and advice during the proposal presentation and the candidature defence.

Last but not least, I want to express my thanks to all my study-mates; we cooperated with each other and learnt from each other. My appreciations will also go with everyone who has helped me directly or inderectly.

TABLE OF CONTENTS

PAGE

LIST OF CONCORDANCES

LIST OF FIGURES

LIST OF TABLES

LIST OF TEXTS

LIST OF ABBREVIATIONS

CLAWS	Constituent Likelihood Automatic Word-tagging System
CMC	Computer-Mediated Communication
EFL	English as a Foreign Language
ENL	English as a Native Language
ESL	English as a Second Language
ICQ	I Seek You
ICT	Information and Communication Technology
IRC	Internet Relay Chat
L2	Second Language
MUD	Multi-User Dungeon
PR	PageRank
POS	Part of Speech
UCREL	University Centre for Computer Corpus Research on Language
USAS	UCREL Semantic Analysis System

CHAPTER 1

INTRODUCTION

1.1 Introduction

This dissertation seeks to explore the content of female and male adults' weblogs in an ESL context, i.e. Malaysia, by using a corpus-driven approach. For this purpose, two sub-corpora in English have been built: a sub-corpus consisting of female weblogs (41, 585 words) and a sub-corpus of male weblogs (39, 222 words). In this chapter, the background of the study is introduced in Section 1.2, followed by statement of the problem in Section 1.3. The objectives and research questions are stated in Section 1.4. Next, five key terms relevant to the thesis are defined in Section 1.5. The chapter ends with a brief organisation of the whole thesis in Section 1.6.

1.2 Background of Study

Contextualization is used as one of the most important cognitive strategies for second language learning (Zare, 2012). Bahrani and Tam (2012) indicate that many research studies have considered the use of interaction and technology in ESL context to facilitate L2 learning (Clifford, 1998; Egan, 1999; Pemberton, Fallahkhair, and Masthoff, 2004). In the present study, the ESL context would reflect the language preference among Malaysian bloggers who write English blogs in Malaysia.

Even if semantic fields have been explored in depth, in the field of structural linguistics, computational approaches for semantic fields have been proposed quite recently by introducing the concept of 'Semantic Domains' (Magnini, Strapparava, Pezzulo, and Gliozzo, 2002). According to Gliozzo (2005: 37), 'Semantic Domains' are "clusters of terms and texts that exhibit a high level of lexical coherence, i.e. the property of domain-specific words to co-occur together in texts. The concept of 'Semantic Domain' extends the concept of 'Semantic Field' from a lexical level, in which it identifies a set of domain related lexical concepts, to a textual level, in which it identifies a class of similar documents. The founding idea is the lexical coherence assumption that has to be presupposed to guarantee the existence of 'Semantic Domains' in corpora".

Gliozzo (2006) establishes a Word Sense Disambiguation (WSD) procedure merely depending on domain message, named Domain Driven Disambiguation (DDD). The work attests that semantic domains have a two-tier function. On one hand, they are capable of depicting word senses by allocating domain tags to word senses in standard corpora. On the other hand, semantic domains are also clusters of texts considering analogical themes at a text level.

Language is a kind of social phenomenon and a part of the culture. (Risager, 2005; Cakir, 2006). It is also a reflection of culture and society (Xia, 2012). So it is a natural

phenomenon that the male and female social status inequality is reflected in language (Lei, 2006). The relation between gender and language has become one of the hot topics in the branch of sociolinguistics since early 1970s (Sunderland, 2006; Wodak, 2009). Then an explosion of related research was carried out in many separate aspects (Ning, Dai, and Zhang, 2010). Linguists have explored the gender differences in pronunciation (Byrd, 1992; Simpson, 2003), intonation (Brend, 1975; McConnell-Ginet, 1978; Daly and Warren, 2001; Lakoff, 2004), vocabulary (Andersson, Gauding, Graca, Holm, Öhlin, Marklund, and Ericsson, 2011; Llach and Gallogo, 2012) and discourse style (Davies, 2003; Jones and Myhill, 2007) from the perspective of sociolinguistics research, and analysed the latest reasons of these differences along with the development and changes of these differences.

From the early 1960s, people in the United States have witnessed the occurrence of the Internet (Cohen-Almagor, 2011). In the era of information, synthesizing the rise of networks with the demand for more convenient communication, “Blog”, a blending of the words web and log, which is “a web page that contains a running log of commentaries, multimedia, and hyperlinks” emerges at the right moment (Armstrong and Retterer, 2008: 234). In the initial phase, original webpages solely showcase texts and photographs uploaded by developers of those websites. To some extent, this mode of operation confines potential website writers to upload their works freely owing to the defective computer skills.

The creation of weblogs inaugurated numerous opportunities for language learners to “express and share their ideas with the boundless internet community” (Nepomuceno, 2011: 92). Weblogs furtively joined the World Wide Web in the late 1990s, but owing to its indubitable pervasiveness among individual blog authors who are eager to share their ideas online, web designers set about developing brand-new blog websites which are more apt to posting and keeping.

This innovation further contributed to the rapid growth of fanaticism in the blogosphere. “In a recent survey, the Pew Internet and American Life Project estimated that some 12 million Americans published their own blogs and 57 million read blogs” (Lenhart and Fox, 2006: 2). According to Lenhart and Fox (2006: 2), this tool, namely, computer-mediated communication (i.e. CMC) “has captured the attention of the nation with articles about blogging appearing in The New York Times (Selingo, 2004), The Washington Post (Kinzie, 2005; Walker, 2004), The Chronicle of Higher Education (Carlson, 2003; Glenn, 2003; Turkle, 2004) as well as in journals devoted to technology such as T.H.E. Journal (Ferdig and Trammell, 2004) and Syllabus/Campus Technology (Long, 2002; Roberts, 2003)”.

Language learners and educators are interested in the use of weblogs because they offer various characteristics: interactivity, accessibility, interest and relevance (Izquierdo and Reyes, 2009; Moon and Lim, 2013). The focus of foreign language

learning has shifted from ‘L2 learning’ to ‘L2 use’ and from ‘classroom context’ to ‘naturalistic settings’ during the last decade (Wang and Vasquez, 2012: 209). The weblog corpus in the present study provides online language used outside any ESL classroom.

The internet is capable of promptly bringing together people from diverse races and cultures worldwide to share life experiences and thoughts with each other in real-time which may be an explanation for its pervasiveness. Other than this, weblog sites also have been designed to be “customizable”, permitting the users to amend the image of their page, such as pictures, themes, colours and so forth. These characteristics of blogging provide the researcher with ideas as to the usefulness of blogs in online communication. Taking gender as the bedrock, the thesis aims to investigate the online language of female and male in the blogosphere.

1.3 Statement of the Problem

While previous linguistic analysis of gender focused on data from daily life conversations and paper-based written materials (Eckert and McConnell-Ginet, 2003; Holmes and Stubbe, 2003; Walsh, 2001), recent studies began to investigate computer-mediated-communication (CMC) forms such as Internet Relay Chat (IRC)[1], I

[1] “Internet Relay Chat (IRC) is a protocol for real-time Internet text messaging (chat) or synchronous conferencing. It is mainly designed for group communication in discussion forums, called channels, but

Seek You (ICQ)[2], or Multi-User Dungeon (MUD)[3], and emails. Owing to the proliferation of public diaries, weblogs, which provide another important domain to explore gender differences, are gaining an increasing amount of attention from linguists (Herring, Kouper, Scheidt, and Wright, 2004; Huffaker and Calvert, 2005; Herring and Paolillo, 2006). So far, most of the existing research has focused on the writing styles, strategies and exploration of identity of the bloggers using a qualitative approach (Thayalan and Noor, 2010; Nizam, Yee, Yussof, Belongkikit, and Ibrahim, 2011; Amir, Abidin, Darus, and Ismail, 2012). However, there has been a lack of studies examining gender comparisons in the content of weblogs on a larger scale using corpus approaches. Using corpus approaches will allow for a more systematic and statistical analysis of naturally occurring data and objective verification of results (Leech, 1992).

For years, the English language has been taught as a second language in all the national schools of Malaysia. Moreover, in the light of the globalization trend, English has been instilled as a compulsory course in curriculum for all the Malaysian national schools (Yunus, Kiing, and Salehi, 2013). Ever since it entered Malaysia in 1998,

---

also allows one-to-one communication via private message as well as chat and data transfer, including file sharing (source: http://en.wikipedia.org/wiki/IRC).

[2] I Seek You (ICQ) is an instant messaging computer program that was first developed and popularized by the Israeli company Mirabilis, then bought by America Online, and since April 2010 owned by Mail.ru Group (source: http://en.wikipedia.org/wiki/ICQ).

[3] A MUD (originally Multi-User Dungeon, with later variants Multi-User Dimension and Multi-User Domain) is a multiplayer real-time virtual world, usually text-based. MUDs combine elements of role-playing games, hack and slash, player versus player, interactive fiction, and online chat (source: http://en.wikipedia.org/wiki/MUD).

weblog ushered a new epoch for Malaysian ESL learners (Hopkins, 2010). Since then, pedagogical research of ESL blogging in writing class has gained interest amongst Malaysian scholars (Mah and Er, 2009; Mah and Liaw, 2011; Gedera, 2012; Kaur, Ganapathy, and Sidhu, 2012; Yunus et al., 2013; Ubaidullah, Mahadi, and Ching, 2013). However, there is little literature from a sociolinguistics perspective, let alone gender studies on weblogs.

Although the pioneering work of Malaysian blog activism sought democracy in politics (Tang, 2005; Smeltzer, 2008; Ahmad, Mohamad, Hassan, Pawanteh, Ahmad and Aziz, 2011), the changes of the research foci came into being gradually. Recently, some researchers started to choose gender as a significant variable when investigating Malaysian blogosphere (Thayalan and Noor, 2010; Nizam et al., 2011; Amir et al., 2012; Mustapha and Wang, 2012).

Amir et al. (2012) investigated four teenage bloggers (two females and two males) who ran their blogs in www.blogmalaysia.com for more than one year with at least 50 postings. The selected blogs span over three months and it was found that female teenagers used more intensifiers, hedging, empty adjectives, tag questions and adverbs. Though the findings were consistent with that of Lakoff's (1975) who stated that females used more intensifiers, hedging, empty adjectives, tag questions and adverbs than the males, the researchers addressed an unequal number of postings for each

gender group. Furthermore, some empirical studies attempting to draw profiles for bloggers (Herring, Kouper, Scheidt, and Wright, 2004) discovered that teenage female bloggers constituted the majority of bloggers.

Insofar there are few studies focusing on studying adult bloggers' behaviors (Bidin and Mustaffa, 2012) let alone studies related to the adult bloggers who write English weblogs. Moreover, there is a lack of literature in the gender issues in Malaysian blogosphere. In order to fill the gap, the present research study is designed to investigate significant semantic domains and parts of speech in the content of two hundred English weblogs (one hundred female blogs and one hundred male blogs) written by Malaysian adults using a corpus-driven approach.

1.4 Objectives and Research Questions

This thesis explores the content of Malaysian male and female weblogs written in English using a software tool for corpus analysis and comparison, i.e. Wmatrix (Rayson, 2013). The tool compares the male and female weblog sub-corpora and identifies key semantic domains and key parts of speech which are significantly used by each gender group allowing comparisons to be done at semantic and part of speech levels.

The thesis aims to address the following three research questions:

- **Research Question One:** What are the key semantic domains in the female weblogs sub-corpus and male weblogs sub-corpus?
- **Research Question Two:** What are the key parts of speech in the female weblogs sub-corpus and male weblogs sub-corpus?
- **Research Questions Three**: How do the differences between the use of key semantic domains and key parts of speech relate to socio-cultural factors?

The first research question is answered through the application of a corpus tool, i.e. Wmatrix, to a specialised corpus of Malaysian adults' weblogs to analyse the semantic domains used by the females compared to males and males compared to females. The second research question is parallel to the first. The same parameters as mentioned above, for Research Question One, are used to analyse the key parts of speech in the corpora. Wmatrix is used to generate word lists and frequencies of the significant semantic domains (tagged by USAS) and key parts of speech (tagged by CLAWS) tagged in each weblog.

Based on the algorithm of Wmatrix, the female sub-corpus is compared with the male sub-corpus. Then, the log-likelihood value determines the key semantic domain and part of speech in the female sub-corpus. Subsequently, the same algorithm is used to

compare the male sub-corpus with the female sub-corpus. Therefore, Research Questions One and Two are further divided into four sub-questions.

Research Question One

i) What are the key semantic domains in the female weblog sub-corpus in relation to the male weblog sub-corpus?

ii) What are the key semantic domains in the male weblog sub-corpus in relation to the female weblog sub-corpus?

Research Question Two

i) What are the key parts of speech in the female weblog sub-corpus in relation to the male weblog sub-corpus?

ii) What are the key parts of speech in the male weblog sub-corpus in relation to the female weblog sub-corpus?

To address Research Question Three, 'Difference Theory' and 'Dominance Theory' are adopted to explain the social-cultural factors that lead to these differences.

1.5 Definitions of Terms

Referring to the key words in the abstract, the researcher defines five related terms as the fundament. Each of these definitions has been exhaustively expatiated with a significant definition which is also adopted in this dissertation.

1.5.1 Corpus

A corpus is a collection of texts used for linguistic analysis through means of computerised technology. As defined by Leech (1997: 1), a corpus is "a body of language material which exists in electronic form and is designed to represent a particular language or language variety". An important question that arises in defining a corpus is the representativeness of the sample (Tognini-Bonelli, 2001). According to McEnery and Wilson (2001: 64), "In discussing the ways of achieving the maximal degree [of] representativeness, it should first be emphasised once again that in producing a corpus we are dealing with a sample of a much larger population". Therefore, the corpus in this study represents adults' weblogs produced by a small sample population of ESL speakers living in Malaysia.

1.5.2 Semantic Domains

In the field of computational linguistics, semantic domains are regarded as a comparatively new topic, even if their essential hypotheses are enlightened from a perennial research orientation in structural linguistics, i.e., the theory of semantic fields (Lyons, 1977). They can be automatically verified by utilizing a lexical coherence quality revealed by texts in whichever language, and can be advantageously employed to construct a semantic network to define a computational lexicon.

1.5.3 Gender

The British sociologist, Giddens, (1989: 158) defines gender as "the psychological, social and cultural differences between males and females". At the early stage of research, the word "sex" is used to refer to both biological and psychological sex. But with the advance of feminism movement in the 1960s and 1970s, "gender" is borrowed by feminist linguists to describe the categories which are structured by social attributes on the basis of sex. Fausto-Sterling (2000) sums up the understanding of 'gender' as follows:

> Labelling someone a man or a woman is a social decision. We may use scientific knowledge to help us make the decision, but only our beliefs about gender - not science. (p. 3)

In this sense, the criterion of judging whether a person is male or female is not only owing to his or her inborn genital organ but also the typical features of his or her apparel, utterance and behaviours which are evaluated in the social categories belong to either of male or female.

1.5.4 Computer Mediated Communication (CMC)

Computer-mediated communication (CMC), which normally appears in a written genre, is an approach of exchanging information with others via the computer. The printed text is transmitted through "electronic equipment" to the computer (Walther, 1992). From the last two decades of twentieth century, various types of CMC has

mushroomed all over the world, such as online-chat rooms, e-mails, weblogs, electronic bulletin boards, etc.

The unprecedented boom of CMC also impacts people's utterance (Chesebro and Bonsall, 1989). The superiorities provided to CMC utilizers not only include exchanging messages regardless of the limitations of space or time, facilitating the interflow among interlocutors, but also inaugurate a brand-new style of communication (Widjaja, 1997). The imperfection in social presence, social contexts and visual channels signal the creativity of a neo-conversational modality. In consequence, the styles in which people interact in the CMC sphere may differ from those of face-to-face conversations. Face-to-face communications occur in collaborative communities constantly managed by mutual correction and adjustment (Galimberti, 1994), whereas CMC occurs in a much less collaborative community owing to its peculiarity (Brennan, 1991).

1.5.5 Weblogs

Weblog is a public or personal website that permits people to post information or share thoughts. In the early era of weblogs, Bausch, Haughey, and Hourihan (2002: 7), while describing the format, define weblogs as "pages consisting of several posts or distinct chunks of information per page, usually arranged in reverse chronology from the most recent post at the top of the page to the oldest post at the bottom." Herring et al.

(2004: 101) also define weblogs as “frequently modified web pages in which dated entries are listed in reverse chronological sequence.”

1.6 Organisation of the Thesis

The rest of this thesis is organised as follows: Chapter 2 presents an overview review of relevant literature in relation to the research. In Chapter 3, the theoretical framework and research methodology of the study are described. Chapter 4 presents the analysis of the data using the software, Wmatrix, and its interpretations. Chapter 5 concludes the thesis by providing a summary of the findings of the research. It also discusses some implications, as well as its limitations, and provides suggestions for future research.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

As introduced in Chapter One, this thesis examines the content of weblogs produced by female and male adults in Malaysia, using a corpus-driven approach. In this regard, this chapter provides an elaborate review of key issues related to the study. As well as providing background context, the aim of this chapter is to position the present study in relation to existing research, noting gaps where they appear, and reviewing such research from a critical perspective.

Firstly, this chapter takes a broad perspective by reviewing researches related to language and gender as presented in Section 2.2. It then examines two branches of CMC and how gender affects the written form of CMC (Section 2.3). Next, the chapter provides an overview of weblogs and reviews relevant studies on influences of gender on weblogs (Section 2.4). Finally, as the methodological framework of this study is corpus linguistics and content analysis, the researcher includes literature in the area of gender differences in weblogs from a corpus perspective (Section 2.5) and via content analysis approach (Section 2.6).

2.2 Language and Gender

Simone De Beauvoir (1989: 295) claims “One is not born, but rather becomes, a

woman.” Corresponding to this, one is not born, but rather becomes, a man. Beauvoir's claim is important because it is among the first few statements in modern feminism to draw attention to a “woman” as a social, rather than a natural, category of being (Ritzer, 2004: 304).

In the broadest sense, gender has been employed by social theorists to denote a distinction between the biological categories of female and male and the socially constructed categories of woman and man (or girl and boy). Gender is revealed to involve the management of situated conduct in adherence with normative conceptions of masculinity and femininity. Gender is thus seen as a highly significant dimension for understanding how the body becomes a social fact (Ritzer, 2004: 304).

However, gender is not merely regarded as a psychological attribute. It involves people’s sexuality and must always be comprehended “in the context of particular, and changing, social relations between women and men” (Kassam, 1996: 112).

The researcher grounded her reviewing from the relation between language and gender, followed by the study of gender related to spoken (Section 2.2) and written language respectively (Section 2.3).

2.2.1 The Relations between Language and Gender

According to Graddol and Swann (1989), there are three ways language and gender can be associated. First, language reflects gender divisions. Coates (1986) holds the view that linguistic differences are reflections of social differences, and as long as society views men and women as different and unequal, then differences in the language of men and women will persist.

Second, language creates gender divisions. Spender (1980) underlines the significance of language in the dimension of human existence.

> Language helps form the limits of our reality. It is our means of ordering, classifying and manipulating the world ... Having learnt the language of a patriarchal society, and we have also learnt to classify and manage the world in accordance with patriarchal order and to preclude many possibilities for alternative ways of making sense of the word. (p. 3).

The second position suggests that language does not function simply as a mirror of society, but suggests that our individual lives and personalities are shaped by our language and by the discourse we engage in. The social divisions and inequalities are actually created through sexist linguistic behaviour (Gallardo, 2001).

And third, a view that argues that "both processes (i.e. language reflects and creates gender divisions) apply, and that any full account of language and gender must explore the tension and interplay between the two" (Graddol and Swann, 1989: 9).

Taking this view into consideration, we may infer that language might, thus, reflect and create gender divisions. As Graddol and Swann (1989: 10) explain, “linguistic and social practices are mutually supportive”.

2.2.2 Theories and Researches on Gender and Spoken Language

Gender and language, as one of sociolinguists’ most interested research subjects have developed over time. Robin Lakoff’s theories on women’s language suggest that most women use a language style that promotes diffidence, shyness, and lower self-confidence, resulting in a lack of commitment or strong opinion. There are three devices for women’s language strategies. One device is euphemism, where a person uses words such as “fudge” or “heck” instead of profanity. Another device is the use of tag questions and hedges, such as “this weather is terrible, isn’t it?” One more device is indirection when there is a reluctance to commit to something, for instance, “Well, I’ve got a doctor’s appointment around that time.” Finally, for Lakoff, women’s language represents an overall conventional politeness (Eckert and McConnell-Ginet, 2003: 158).

In Lakoff’s (1975) article on gender, language, and power, she argues that women’s conversational strategies are affected by their oppression under patriarchy. What Lakoff calls “women’s language” is marked by powerlessness in the forms of “super politeness,” qualifiers, exaggerations, and tag questions. These weak conversational forms provide “diagnostic evidence” for the inequality between men and women. Lakoff further

theorizes that women are socialized to speak in ways and are perceived as weak which, in turn, reproduce their oppression. Thus, Lakoff argues that there are important differences between men's and women's speeches.

Fishman's (1978; 1983) work looks at men's power over women in everyday interactions. Fishman argues that formal features that characterize women's speech (e.g., asking questions) seek to insure response. In contrast, men's speech is marked by features (e.g., statements) that do little to insure further talk. However, interlocutors are more likely to "orient to" topics that men raise; "the definition of what is appropriate or inappropriate conversation becomes the man's choice" (Fishman, 1983: 98). In other words, men control talk. Moreover, women generally work harder than men to facilitate conversation by taking on what Fishman (1978: 405) terms as the 'shitwork' in the interaction (Byrne, 2005). Thus, Fishman uses the binary facilitative/controlling to describe women's and men's speech, respectively.

Subsequent researches, however, have criticized Lakoff's (1975) and Fishman's (1978; 1983) claims. Some studies suggest that interactional context affects the extent to which men's and women's speech can be distinguished (Cameron, McAlinden, and O'Leary, 1989; Graddol and Swann, 1989; Swann, 1989; James and Drakich, 1993; Freed, 1996).

Lakoff's and Fishman's critics argue that an utterance's function cannot be placed into a single category or be known in advance. Tag questions, Cameron et al. (1989: 85) argue, are "characterized by complex multi-functionality and diversity of meaning". Graddol and Swann (1989: 89) argue that discrepancies between men's and women's speech are "differences of degree" and that lists of women's and men's speech styles constitute a "gross over-simplification". Thus, representing male and female language as a binary and using other dichotomous categories to describe language overlook complexities in actual speech.

Speech was also described using dualistic categories in later researches. In reaction to Lakoff's and Fishman's "dominance" perspectives, feminist linguistic scholarship began to attribute discrepancies in male and female speech patterns to "cultural differences" (Cameron, 1992; 1996). Tannen's studies (1990) exemplified the "cultural difference" approach (Cameron, 1996). While Tannen did not deny the existence of male dominance in society, her analysis downplays patriarchy's role in producing linguistic differences. Tannen (1990: 77) argued that "because boys and girls grow up in what are essentially different cultures…talk between women and men is cross-cultural communication". Cultural differences explain why women use conversation to connect with others and for "negotiating relationships". For men, conversation is for "holding center stage" and for "get [ting] and keep [ing] attention"; talk maintains independence

and status. Tannen (1990) called women's talk 'rapport-talk' and men's talk 'report-talk'. Thus, speech functions are constructed using the binary rapport/report.

Tannen (1990) points out that gender discrimination in language habit results in the different expectation and demand coming from society. Tannen (1990: 24-25) claims that men have strong status sense and show the tendency to competition during conversation while women appreciate keeping good relationship with the counterparts. Therefore, women who normally use euphemisms, tend to agree with others' opinions, and seldom present suggestion. In the view of women, conversation process is the only feeling which exchanges without considering much of conversation contents. They try to establish and strengthen the relationship. Rodino (1997) summarizes the overall critiques of gender and language research as follows:

> Asking questions about how speech patterns of men and women differ creates problems for feminist linguists. Describing speech features as "male" and "female" overlooks the extent to which context influences the ways that utterances operate. In addition, contrasting "male" and "female" language reifies differences between men and women. Such distinctions help rationalize women's oppression. The relationship between language and gender helps to expose biological essentialism, the binary gender system, and patriarchy as cultural constructions (p. 0).

2.2.3 Studies in Written Language and Gender

So far, the mainstream of gender-bound language research pertains to speech rather than texts in written formats (Schultz, 2013). Jones and Myhill (2007) conducted a

two-year research project in which gender and linguistic competence of writing were considered two core variables. Stratifying 120 girls and 120 boys who were all secondary-aged pupils by year group and text type, they analysed gender differences of linguistic characteristics at text as well as sentence level in their 718 excerpts in the category of narrative and argument. The result indicated that more significant differences appeared in the latter other than the former in terms of 'text length', 'paragraphing', 'linking devices' and 'organization' (Jones and Myhill, 2007: 466). Furthermore, the quantitative comparison manifested that girls tended to use shorter sentences and finite verbs whereas boys' writings presented a higher satisfactory in paragraphing and organization (Jones and Myhill, 2007: 472) providing a contrast to the conventional judgment in previous studies that girls are better writers.

Still in ESL and EFL classrooms, other researches also notified gender differences in linguistic features of learners' writing. Regarding writing quantity, female students consistently wrote more than male students (Lee, 1996; Chiu, 2008). In terms of writing quality, Kann (2001) illustrated that female students' performances were far better than that of male students in the aspects of grammar, content, diction and organization.

Applying the same analytic framework, Chiu (2008) analysed 280 paragraphs written by seventy non-English major undergraduates (thirty-five females and thirty-five males) in Taiwan. Each participant had to write four English paragraphs

including cause-effect paragraphs, descriptive paragraphs, comparison paragraphs and narrative paragraphs in sequence. Taking mechanics as an additional criterion, different writing performances were revealed between the two genders. Female students performed better than male students in terms of narrative and descriptive paragraphs. Nevertheless, male students wrote slightly better than the female ones in the comparison and cause-effect paragraphs. These results were in accordance with Goldberg and Roswell's (2002) research in which girls were observed to organize their writings by their own experiences and observations whereas boys are adept at reasoning and debating.

Hasan and Khammat (2011) analysed sixty e-mails posted by sixty Iraqi freshmen who majored in English and investigated the gender differences in the usages of sentence types and modifiers. Findings indicated significant differences in the former whereas slight differences were located in the latter. Female students tended to use more simple sentences whilst male students preferred to use more complex, compound, and complex-compound sentences. With regards to the modifiers, female students slightly surpassed the male ones in the total number of modifiers but differences located in the proportion of specific types. The female students tended to express uncertainty and doubt with modifiers such as *perhaps*, *maybe* and *probably*. On the contrary, the male students made more references to certainty by using modifiers including *certainly*, *of course*, *surely*, *in fact* and *naturally*.

2.3 CMC and Gender

The language on the Internet represents a new type of discourse that is shaped by the creativity and innovation of its community (Crystal, 2001). In 1992, Susan Herring published her first paper related to CMC and since then, she has made enormous contribution to CMC research field. In 1993, using electronically-distributed questionnaires, Herring found that women were more likely than men to react passively to aggression in online interaction, including falling silent and dropping out of conversation groups.

Since the appearance of CMC, many scholars regarded it as a gender-free community. Herring (2001: 623) summarized CMC as a medium devoid of physical cues and renders it possible that online communication can be equal between men and women. This is totally different from the traditionally male-dominated face-to-face communication. Besides, the Internet can make a connection between all users who are even geographically dispersed, thus the subordinated group can be empowered the equal right to pursue their interests through this online community. Lastly, World Wide Web can serve as a medium for women's self-publicity and lucrative activity.

But in the following research, Herring constantly challenged her previously raised opinions. For instance, she opined that the number of female Netizens was far smaller than their male counterparts and there was certain discrimination in the online world

where some activities were not open to females. Herring also believed that CMC was dominated by males and therefore not easy to use for females; as a result, females hesitated using CMC (Herring, 1992; 1993).

In 1993, Herring mentioned that males and females showed obvious distinctions concerning the preferred online topics with males being fond of talking about politics and information while females were keen on families and personal affairs.

In addition, it was found that 68% of male participants showed hostility in attitude, kept certain distance from others and often put an emphasis on self. In contrast, female participants were more polite in their language use. They tended to raise more questions and waited for others to provide a solution for many issues. Women were also good at expressing their emotions and opinion with a purpose to get some interaction and support (Herring, 1994).

2.3.1 Synchronous CMC (SCMC) and Asynchronous CMC (ACMC)

Harasim (1990) classified CMC into two sub-categories, as follows:

(1) Synchronous CMC: communicating by means of computers and other digital technologies at the same time, and

(2) Asynchronous CMC: communicating by means of computers at different times.

The dominance of asynchronous CMC enabled people to reply more rationally by providing enough time to comprehend the information. In the present study, the researcher examined one typical kind of asynchronous CMC, i.e. weblog. In order to explore gender differences and similarities in different CMC modes, the researcher first drew a distinction between synchronous CMC (SCMC) which happens in real time, such as instant messaging- and asynchronous CMC (ACMC) - which serves as a delayed-time interaction such as weblog or email. According to Abrams (2003), several differences and similarities between these two types of electronic communication are evident (Table 2.1).

Table 2.1 Differences and Similarities in Synchronous CMC and Asynchronous CMC

	Synchronous CMC	**Asynchronous CMC**
Differences	1. Relatively immediate responses 2. Use of outside resources cumbersome 3. Social immediacy of interlocutors	1. Extended planning, encoding, decoding time 2. Use of outside resources not limited 3. Interactants do not "immediately" present
Similarities	1. Extensive learner-to-learner (or learner-learner-teacher) negotiation of meaning 2. More "talk" time per learner than oral classroom communication 3. Increased amount of output results in richer and more diverse lexicon 4. Written code 5. Register between those of written and oral styles of communication	

Comparisons about gender differences in synchronous CMC and asynchronous CMC are mainly carried out in the settings of classrooms (Chou, 2002). After conducting a content analysis of the transcripts from a designed course, Chou (2002) found that female undergraduates contributed more to social-emotional-oriented interaction than the male counterparts in both synchronous CMC (online seminars during the designed course) and asynchronous CMC (discussions after the designed course). They also produced more messages in the synchronous CMC than the male students.

As for gender differences in the two communication modes respectively, scholars probed into not only the application in language learning but also the discourse online by comparing with face-to-face communication (Dalampan, 2006; Yip, 2006; Sun, 2008).

2.3.2 Researches Related to CMC and Gender

The Internet is widely regarded as a democratic place that levels the playing field between gender and socio-economic power (Herring, 2001). However, early research into gender and CMC suggests that power struggles and gender bias found in the real-world are replicated online. There are, however, disagreements on gender differences in CMC. Some researchers argue that females are disadvantaged by

socio-cultural reproductions, a lack of access to technology or even inferior technological fluency (Gunn and McSporran, 2003).

The first research on gender and CMC was published in the late 1980s by Smith and Balka (1988). This study reported the employment of the Internet to connect geographically-dispersed women and facilitate feminist activism (Herring, 1992).

Graddol and Swann (1989) conducted a study in a university and found that men and women seemed to have been empowered equal rights through employment of anonymous computer conferencing system. This result formed an obvious contrast to the traditional pattern of men's domination in a similar situation. For the most part, however, early CMC research did not discuss gender or control for it in experimental studies (Herring, 1992).

In 1990s, the study of gender and CMC was flourishing due to a proliferation in women's participation in the CMC sphere, which had been previously regarded as a domain dominated by male. In contrast to optimism of the 1980s, the study in this period tended to question gender quality in cyberspace (Herring, 2000).

Selfe and Paul (1991) conducted a study and found that those who had a high status in social life tended to play a dominant role in the online interaction. This study

denied the traditionally claimed cyberspace as a status free sphere. Soon after, many researchers began reporting the use of more aggressive tactics by men in online discussions, some of which explicitly targeted female participants (Herring, 1992; Sutton, 1994). Around the same time, reports began to surface in the popular press of women on the Internet being the targets of male intimidation, harassment and sexual deception (Dibbell, 1993; Van Gelder, 1990).

Why do there exist so many distinctions between men and women in this online community where physical cues can be eliminated? Maclaran and Miriam (2002) claimed that there was no gender discrimination for computer itself, yet technology had gender inclination. As a result, there are gender differences in their topic choices. Finally, they concluded it was social reality that engendered such distinctions.

Hock (1999) shared similar opinion with Maclaran. He pointed out that Internet activity was as a matter of fact incorporated in the social structure and social culture and therefore, served as reflections of social reality. Males and females were never on the same way.

Herring (1994) conducted an analysis about the two websites “Linguist List” and “Politics” and found that there were two reasons which could account for the online gender difference phenomenon. One was style distinction, with the other being their

distinctive interactional ethics. Males tended to be self-centred. The typical male style was to belittle others while raising self-status. Herring (ibid: 4) characterized the female conversation style as "supportiveness" and "attenuation." Attenuation refers to the notion that women are not inclined to intensify contradictions. They are inclined to hold a hesitated attitude towards contrasting opinions or raise their own opinion in a way of suggestion. They prefer to compliment, give support and kind suggestions to others.

2.4 Weblog and Gender

According to Herring et al. (2004), weblogs can be categorized into four genres in terms of their purposes as follows:

(a) Filter: The contents of filter weblogs are external to web-bloggers, and include international and national events (Blood, 2002), or links to websites (Winer, 2003), issue-focused weblogs and political weblogs (such as weblogs about Malaysian General Election in the year of 2013).

(b) Diary/Personal journal weblogs: The content of diary weblogs includes the personal feelings and thoughts of the author (Blood, 2002). They are culturally and socially structured, and reflect ideas related to our interaction within a certain cultural community and the cultural influences the community brings to us (Huff, 1996).

(c) K-log (knowledge log): A K-log is created as a space for knowledge-sharing (Festa, 2003) and is usually technological in essence (Herring and Paolillo, 2006).

(d) Mixed purpose: A weblog of mixed purposes, which has more than one

function, i.e. diaries, k-logs, and weblogs.

(e) Other: Some weblogs do not have any function mentioned in the previous four categories. For example, a weblog may be a space for song writers to create their lyrics (Herring et al., 2004).

On the whole, personal journal weblogs turn out to be the most prevalent which is capable to integrate personal thoughts with characteristics, like the use of different colours in different settings to interact with audience. A weblog also establishes a bond between bloggers and readers in which readers can comment whereas bloggers can reply whenever and wherever (Scheidt, 2006).

Weblogs differ from traditional web pages in two aspects. On one hand, weblogs require a stricter format than personal web pages. On the other hand, it is extremely convenient for bloggers to post their real-time writings on their weblogs with simple techniques (Huffaker, 2004). Furthermore, weblogs also encompass a "comment area", namely, "comments". In this section, readers can share their replies among themselves on certain articles with people who subsequently go through the lines (Winer, 2003).

2.4.1 Gender Issues in Weblog

Gender differences in weblogs have been explored in many aspects, such as topics (Cameron, 1997; Li, 2005), linguistic features (Huffaker, 2004; Kavanagh, 2010; Tossell,

Kortum, Shepard, Barg-Walkow, Rahmati, and Zhong, 2012; Barjesteh, Vaseghi, Hashemi, Pourshahian, and Kalajahi, 2012), and bloggers' personality (Nowson and Oberlander, 2006, 2007; Guadagno, Okdie and Eno, 2008; Wang, Lin and Liao, 2010; Yarkoni, 2010).

Li (2005) conducted a statistical analysis of 280 active adults' blogs randomly selected from http://blo.gs/. The results partly supported the hypothesis that men preferred blogging about "external topics", insomuch as male only outnumbered female bloggers in technology and science, business and politics and politicians. However, female exceeded male bloggers in all the four "internal topics", namely, "interests and hobbies, family and friends, own creative work and personal experience" (Li, 2005: 72).

By analysing eighty blogs (forty Japanese blogs and forty American blogs) categorized into *Travel*, *Sport*, *Family* and *General*, Kavanagh (2010) stated that Japanese bloggers used much more emoticons than American bloggers. In the Japanese corpora, female bloggers used most emoticons in travel blogs, whereas male bloggers adopted emoticons most frequently in family blogs. For both genders, sports blogs were found to produce the least number of emoticons.

Regarding gender differences in disclosure of personal information, Huffaker (2004) indicated that male bloggers preferred to uncover their locations, while female

bloggers were in favour of a link to a personal webpage after analysing seventy teen weblogs. But he did not find significant differences between the two genders in the choice of online name and avatar selection in contrary to previous studies.

Others make thorough inquiries into various types of weblogs. Political blogs (Fallon, Williamson, and Pack, 2011; Smith, 2011), travel blogs (Pan, MacLaurin, and Crotts, 2006; Wenger, 2008), medical blogs (McNamara, 2007; Tseng, 2008), religious blogs (Pollock, Okdie, and Guadagno, 2011) and such have been investigated in depth by scholars worldwide.

2.4.2 Weblog and Gender in Malaysian Blogosphere

Mustapha and Wong (2012) examined the perception of Malaysian bloggers generally and the dynamic use of English among the four top-ranked Malaysian blogs[4] by conducting a survey research with thirty final-year TESL undergraduates. Most respondents agreed that blogs did play an important role in their daily lives but the influence was not so remarkable regarding their language use. Meanwhile, they recognized the high proficiency of English usage in the four selected blogs and the typical linguistic strategies, namely, code-switching and code-mixing which made the blogs easier to comprehend.

[4] www.kennysia.com., www.sapiensbryan.com., www.paultan,org., and www.chedet.com (Source: Gaman, 2007)

When gender emerged as an issue, Nizam, Yee, Yussof, Belongkikit, and Ibrahim (2011) investigated the gender disparities and awareness among 229 Malaysian high education institution students via questionnaires. They discovered that the bloggers' own interest shaped their blogging and was in accordance with their gender characteristics affected by subjective norms, personal result expectation and self-expression. Findings also suggested that females were frequent bloggers and tended to share social life experiences, whereas the top three types of males interests were music, information technology, as well as film and TV series.

Differing from previous studies that males are more direct in gender-based interaction, Thayalan and Noor (2010) review comments and feedback messages posted at the researcher's blog by fifty-three students over thirteen months and state that female participants are straightforward in views by "making blatant and sexist remarks and giving blunt advice" (Thayalan and Noor, 2010: 900). They are also observed to be more prevalent with gender-based strategies (such as stereotyping strategy, alliance building strategy etc.) than male participants.

2.5 Corpus Linguistics

Corpus linguistics has been important to debates in linguistics since the 1980s with its divergence from the traditions of a Chomskian approach of intuition and assumed ideas about language to an empirical-based study of real language use following

Firthian and Sinclarian approaches to language. Interest in corpus approaches has contributed towards the 'intuition' versus 'evidence-based' debate, the latter forming the basis of corpus linguistics. Perhaps the most comprehensive explanation for what constitutes a 'corpus' is that given by McEnery and Wilson (2001) who considered a corpus under the following four headings:

1. sampling and representativeness;
2. finite size;
3. machine-readable form; and
4. a standard reference. (p. 29)

This classification indicates the corpus as encompassing a sample of the language variety being represented. The strength of corpus linguistics lies in its providing the researcher with a means of quantifying linguistic features through statistical measures of significance via the application of computer technology, which otherwise, is not normally possible to do manually. Leech (1992) describes corpus linguistics in terms of its application of computer technology. He states that corpus linguistics:

> Defines not just a newly emerging methodology for studying language, but a new research enterprise, and in fact a new philosophical approach to the subject. The computer, as a uniquely powerful technological tool, has made this new kind of linguistics possible. (p. 106)

With this methodology, the application of appropriate measuring tools is required which, as indicated by Sinclair (2004: 189), are "tools of indirect observation, like query languages, concordances, collocates, parsers and aligners".

2.5.1 Corpus-linguistic Research on Gender

A fundamental purpose of the present research is to further corpus research on semantic domains and parts of speech (POS). According to Biber, Conrad and Reppen (1998: 221-222), males prefer to write in an "informational" style whereas females adore an "involved" style. As for gender-featured semantic domains, scholars have explored the gender differences both in spoken language (Johnstone, 1993; Schmid, 2003; Nowson, 2006; Flekova and Gurevych, 2013) and in written texts (Peterson, 2002; Newman, Groom, Handelman and Pennebaker, 2008; Kaur, 2009).

Newman et al. (2008) claim that males express their current concerns by writing about money, leisure, and sports. Conversely, the female interests pertain to relationships (Argamon, Koppel, Fine and Shimoni, 2003) within home, family, and friends which are more personal than the male objective style.

Koppel, Argamon and Shimoni (2002) further classify the gendered-stylistic features into male features ("determiners, numbers, modifiers") and female features ("negation, pronouns, certain prepositions"). Since then, gender differences in POS are

explored by a large number of scholars (Biber, et al., 1998; Koppel et al., 2002; Argamon et al., 2003; Schler, Koppel, Argamon, and Pennebaker, 2006; Newman et al., 2008; Bednarek, 2009; Ali and Aslam, 2012; Yu, 2013; Rangel and Rosso, 2013; Flekova and Gurevych, 2013). Overall, females use more pronouns, interjections, adverbs and verbs while males use more preposition, articles and nouns. A more elaborate review is displayed chronologically in Tables 2.2 and 2.3 respectively.

Table 2.2 Corpus-driven Analysis of Gender Differences in Semantic Domains

Scholar(s)	**Findings**
Flynn (1983)	The male students made few references to women and often wrote on typically "male" topics: *gun control, nuclear power*, or *cars*. In contrast, Women frequently described *accommodation* to the environment rather than rebellion against it.
Johnstone (1993)	In oral narratives, male narrators gave more references to *place* and *time* than female narrators.
Levine & Geldman-Caspar (1996)	Almost 60% of the boys preferred to write about *science-related inventions*, whereas girls' preferences were distributed among the different tasks.
Gambell & Hunter (2000)	Females were more likely to write about *romance*, whereas males tended to write about *heroic actions*.
Peterson (2002)	Domains that were distinctly male were topics surrounding *violence* and *sports*. Domains that were distinctly female dealt with *romance* and *relationships*.
Schmid (2003)	An over-representation in WOMEN was confirmed for the domains *clothing, basic colours, home, food and drink, body and health* as well as *people*. Words and expressions from the domains *work, computing, sports,* and *public affairs* tended to be found more often in MEN than in WOMEN.
Nowson (2006)	Men discussed more impersonal topics, while women preferred those of a more personal nature. Words referencing to *Humans and Family*, *Communication and Hearing* as well as *Emotions* are women' preponderance. They also used more terms relating to *Physical states and functions*, while men talked more about *Money*.

Table 2.2, continued

Table 2.2 Corpus-driven Analysis of Gender Differences in Semantic Domains

Newman et al. (2008)	Women were more likely to make reference to *home, family, friends* and various *emotions*. Men preferred to discuss *current concerns* (e.g. *money*, *leisure*, or *sports*)
Arici (2009)	Although both girls and boys had the subjects of *love* and *education* in common, girls preferred *health, fears* and *books* more, while boys preferred *politics* and *sports* more.
Kaur (2009)	Malaysian boys tended to use words relating to *transport, buildings* and *animals* whereas the Malaysian girls' writing showed significant themes relating to *plants*, *food* and *humanities*.
Flekova & Gurevych (2013)	Men talked more about *computers* whilst women were more likely to talk about *love*.

Table 2.3 Corpus-driven Analysis of Gender Differences in Parts of Speech

Scholar(s)	**Findings**
Mulac & Lundell (1994)	Men tended to use more *numbers* while female writers were more likely to use *progressive verbs* and wrote longer sentences.
Biber et al. (1998)	Female authors used a more 'involved' style, characterized by more *pronouns* and *present-tense verbs*, while male authors tended to use a more 'un-involved' or 'informational' style, characterized by more *nouns* and long words.
Koppel et al. (2002)	A higher percentage of *determiners, numbers*, and *modifiers* characterized the male style, and a higher percentage of *negation, pronouns*, and certain *prepositions* characterized the female style.
Argamon et al. (2003)	*Pronouns* (female marker): *I, you, he* and *she* were significantly more by females in both fiction and non-fiction. But male authors used more plural pronouns in fiction and more male third-person pronouns (*he, him*) in both fiction and non-fiction. *Specifiers* (male marker): In both fiction and non-fiction, male authors used more post-head noun modification with an *of phrase*.
Nowson (2006)	As well as using more *First-person pronouns*, women also talked about other people a great deal more than men with a high frequency of *Third-person pronouns*. However, men were prolific users of *articles*.
Schler et al. (2006)	Female bloggers used more *pronouns* and assent/negation words while male bloggers used more *articles* and *prepositions.*

Table 2.3, continued

Table 2.3 Corpus-driven Analysis of Gender Differences in Parts of Speech

Scholar(s)	**Findings**
Newman et al. (2008)	Women tended to use more *pronouns* (esp. first-person pronouns) and *verbs*, while men commonly used longer words, more *articles* and *numbers*.
Bednarek (2009)	Female characters used more exclamatory emotive *interjections* (1009 occurrences) than male characters (239 occurrences).
Hamdan (2011)	The Arab male novelist preferred starting his paragraphs with nominal sentences that start with *nouns* or *pronouns*. On the contrary, the Arab female novelist tended to use *verbs* to start her paragraphs.
Ali & Aslam (2012)	The Pakistan females used more *adjectives, adverbs, pronouns* and *prepositions* in code mixed SMS than Pakistan males whilst the males used more *nouns, verbs* and *interjections* in code mixed SMS than their female counterparts.
Flekova & Gurevych (2013)	Men tended to use more *articles*, longer words.
Rangel & Rosso (2013)	Women used more *determinates, interjections* and *pronouns* than men while men used more *adjectives, adverbs, prepositions* and *verbs* than women.
Schultz (2013)	Significantly higher frequencies of *pronouns* and *verbs* were noted in female writing. *Numbers* showed up as significant predictors of male writing in this study, too.
Yu (2013)	In congressional speech, male legislators consistently used more *articles, verbs, adverbs, second-person pronouns* and *subjective first-person pronouns*; female legislators consistently used more *nouns, adjectives, third-person pronouns*, and *possessive first-person pronouns*.

2.5.2 Corpus Linguistics in Malaysia

There has been a growing interest recently in corpus linguistics in Malaysia, specifically in areas such as learner corpora (Abd. Samad, Hassan, Mukundan, Kamarudin, Syd Abd. Rahman, Md. Rashid and Vethamani, 2002; Botley, De Alwis,

Metom, and Izza, 2005; Knowles and Zuraidah, 2004; Botley and Dillah, 2007; Botley, Metom and Dillah, 2007), Malay linguistics (Nor Ida, 2005), English for Specific Purposes (Suad, 1999; Afida, 2005) and textbooks (Bahiyah, Mohd, Kesumawati, Yuen, and Azhar, 2008).

These studies have used different methodological approaches to study a variety of linguistic features, not only in English but also in the Malay language. However, these studies have not examined either semantic domain or POS used among adults. This study is, therefore, to examine the corpus in an adult context in an ESL setting.

2.6 Content Analysis

Depicted as "the scientific study of content of communication", content analysis studies the content which refers to the "meanings, contexts and intentions contained in messages" (Prasad, 2008: 173). Acting as a research instrument concentrating on the authentic content and internal characteristics of media, content analysis has long been recognized as an effective approach to the investigation of communication messages. The basic practice of content analysis is to make a conversion of the non-quantitative messages such as speech, written text and images into quantitative data. It is employed to ascertain the presence of certain words, characters, sentences, etc., and to quantify these kinds of data in an objective manner (Berelson, 1952).

2.6.1 Application of Content Analysis

Though content analysis is utilized by researchers from a variety of fields such as communications, political science, social sciences, language studies, psychology, and history, it is most prevalent and widely adopted in the research of mass communication and social science. It has been used broadly to "understand a wide range of themes such as social change, cultural symbols, changing trends in the theoretical content of different disciplines, verification of authorship, changes in the mass media content, nature of news coverage of social issues or social problems such as atrocities against women, dowry harassment, social movements, ascertaining trends in propaganda, election issues as reflected in the mass media content, and so on" (Prasad, 2008: 177).

One of its most important applications has been to study social phenomenon such as prejudice, discrimination or changing cultural symbols in the communication content. Heuer, McClure and Puhl (2011) searched five major news websites for articles entitled "obesity". By scrutinizing the portrayals of 441 individuals, the study stated that "the majority (72%) of overweight and obese individuals depicted in online new photos were stigmatized" (Heuer et al., 2011: 9). This phenomenon not only brought about unfair prejudice to those vulnerable people in social activities but also resulted in severe consequences in their emotion and physical health (Puhl, Moss-Racusin, and Schwartz, 2007).

Another continual adoption of content analysis is to study the changing trends in the theoretical content and methodological approaches by content analysing journal articles of the discipline (Crawford, Pollack, and England, 2006; Chang, Y. H., Chang, C. Y., and Tseng, 2010). Adopting Braun's (2007) "scientometrics method", Chang et al. (2010) conducted a development trends analysis of science education research by analysing 3,039 published articles in four journals during the period from 1990 to 2007. Findings revealed that empirical studies dominate the categories, contrarily, review, position and theoretical papers were rarely presented in those journals. Regarding topics, students' conceptions and conceptual changes enjoyed the predominance in the period of 1990 to 2007 (Chang et al., 2010: 321) although it slightly declined after the year 2000. Instead, topics concerning "students learning context, and social cultural and gender issues" drew more attraction currently (Chang et al., 2010: 323).

One more significant area of its utilization has been the analysis of newspaper content of election coverage and editorial treatment to mould the opinion of voters (Coleman and Wasike, 2004; Mccluskey, 2005; Adhami, Khademian, Almasi, and Rafiei, 2012; McMenamin, Flynn, O'Malley, and Rafter, 2013). Adhami, et al. (2012) analysed the editorial content of two Iranian newspapers, namely, *Mardom Salari* and *Siyasat-e-Ruz*, which provided follow-up reports on the parliamentary election campaign for the Ninth Islamic Consultative Assembly. As *Mardom Salari* supported the reformist alliance whereas *Siyasat-e-Ruz* were in favor of the conservative alliance,

more number of editorial articles were published in the former which indicated that the reformist alliance were more concerned with public participation. *Mardom Salari* was also noticed to "put more editorial articles at the top of the front page to create a sense of agenda setting and better influence the readership" (Adhami, et al., 2012: 10711).

Content analysis has also been used to ascertain trends in the communication content of dailies, weeklies, cartoons, and coverage of development news, political news and crime news (Murty, 2001; Arthur, 2012).

Other important applications of the method were systematic analyses of advertisements in newspapers and magazines to draw useful inference on national culture, as well as media preferences of advertisers (James, John, and Hensel, 1995; Schafferer, 2004; Li, An, and Yang, 2007; Robinson and Callister, 2008; Fowler and Ridout, 2009; Morris and Nichols, 2013). Similarly, television, radio, and movies offer rich sources of material for content analysis. Many scholars have explored changes in women's roles, sexual behaviour and health, and violence by analysing the content of television and movie messages (Davalos, D. B., Davalos, R. A., and Layton, 2007; Saito, 2007; Feng and Karan, 2011).

The above examples throw light on the diversity and range of relevant studies, which applied the approach of content analysis. Though a versatile method, an understanding of these will help in the use of the method effectively.

2.6.2 Content Analysis of Weblogs

In proposing a framework for analysing blogs, Trammell (2004) suggested content analysis as a paramount and effective means for understanding blogs. Several studies have examined content analysis of weblogs in various areas of interest, such as corporate blogs (Cho and Huh, 2010), medical blogs (Kim, 2009), fashion blogs (Halla, 2009; Lövheim, 2011), travel blogs (Pan et al., 2006), sports blogs (Clavio and Eagleman, 2011; Hambrick, Simmons, Greenhalgh, and Greenwell, 2010), and political blogs (Nahon and Hemsley, 2011; Reese, Rutigliano, Hyun, and Jeong, 2007).

Kim (2009) analysed the content of cancer blog posts in order to describe and explore cancer blog contents. The results of the study showed that the majority of the cancer bloggers wrote short postings. The words ‘cancer’, ‘breast’ and `women’ were ranked the top three most frequent words, showing that breast cancer was the most frequently discussed type of cancer.

Pan et al. (2006) analysed forty blogs which represented visitors’ travelling experiences in Charleston. In the content, several clusters of keywords were identified.

It was indicated that “Charleston” was the most prominent cluster followed by the driving experience during the journey. Clusters specifically associated with “plantations” ranked the third owing to its unique significance among local attractions. The kaleidoscopic nature of travel blogs which was to describe travel experiences was manifested from the result.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

The literature review in the previous chapter suggests that hardly any studies have used corpus linguistics to examine Malaysian ESL adults' use of semantic domains and part of speech in their weblogs. Studies have also shown that corpus linguistics has the potential to reveal patterns of semantic domains and part of speech via statistical measures using computer tools. Therefore, this chapter provides a description of the research methodology employed in the present study in order to examine key semantic domains and parts of speech in Malaysian female and male weblogs, and afford possible explanations to support the main findings.

The chapter is divided into four sections. Section 3.2 formulates the theoretical framework and the software tool for data analysis, namely, Wmatrix, is elaborated in Section 3.3. Next, in Section 3.4, a detailed description of the research methodology presented consisting of the introduction to the corpus (Section 3.4.1), data collection procedures (Section 3.4.2) and data analysis procedures (Section 3.4.3).

3.2 Theoretical Framework

This study presents a corpus-driven approach to the study of key semantic domains and POS through concordance analyses in order to identify items of meaning which

reflect weblogs. In the light of Research Questions One and Two, the researcher establishes her theoretical framework on the basis of the corpus-driven approach.

Within corpus linguistics, there are two different approaches: corpus-based and corpus-driven. A corpus-based approach involves using a corpus to explore existing hypotheses or focusing on pre-determined sets of words or categories. A corpus-driven approach (which the current study employs), however, according to Tognini-Bonelli (2001: 84) stems from discovering the data where:

> ...the commitment of the linguist is to the integrity of the data as a whole, and descriptions aim to be comprehensive with respect to corpus evidence. The corpus, therefore, is seen as more than a repository of examples to back pre-existing theories or a probabilistic extension to an already well defined system.

The distinction is perhaps more of a theoretical one, and many studies actually incorporate elements of both techniques. Kaur (2009) set a good example of the use of the corpus-driven approach to Malaysian and British girls' and boys' vocabulary usage in writing. The study used computer analysis of interpretation of concordances to examine sense relations to identify the paradigmatic relations in the investigated corpus. The findings revealed that there are four significant themes in boys' writing, namely, *Sports and games*, *Aggression and adventure*, *Geographical locations*, and *Male fictional and non-fictional characters*. Nevertheless, in the girls' context, *Relationships*, and *Female fictional and non-fictional characters* were the two most popular topics.

While a large part of the present analysis is corpus-driven (e.g. the research focuses on frequent and key words, without determining what they are), there is a corpus-based approach in that the researcher applied pre-determined semantic categories (i.e. USAS) to her data.

3.3 Data Analysis Software Tool

As mentioned in the literature review, a computer-readable corpus can only be processed in plain text without any additional information. Therefore, the corpus was transferred into plain text using notepad before running it through Wmatrix.

3.3.1 Exploring Wmatrix

Wmatrix, a corpus linguistics software, is run in a web-based corpus processing environment so as to analyse the relevant semantic tags for a set of 200 weblogs. Quantitative data was derived for both the male and female weblogs in terms of semantic domains and parts of speech following the UCREL Semantic Analysis System (USAS) established by Paul Rayson in Lanchester University and the Constituent Likelihood Automatic Word-tagging System (CLAWS).

As an efficient corpus analysis and comparison tool, the Wmatrix program "is an online integrated corpus linguistic software basement in which texts can be uploaded and analysed for word frequency profiles and concordances, annotated in terms of

part-of-speech tagged by the CLAWS and word-sense tagged by semantic content (USAS), and statistically compared against standard corpora samplers" (Rayson, 2003; 2013). Utilizing Wmatrix, some research report bloggers' personality (Ooi, Tan, and Chiang, 2007; Nowson and Oberlander, 2007). Moreover, Wmatix is also adopted to analyse metaphor patterns, such as the WAR metaphor in business magazines, the MACHINE and LIVING ORGANISM metaphors in a novel and in business magazines (Koller, Hardie, Rayson, and Semino, 2008; Hardie, Koller, Rayson, and Semino, 2007), along with indirect metaphor, direct metaphor or implicit metaphor in news texts (Krennmayr, 2011).

In terms of blogging, Ooi et al. (2007) conducted a corpus analysis of semantic categories in 50 personal weblogs in Singapore English with Wmatrix. By measuring the variables of age and gender, they found that as each year goes by, female teenagers tended to be less clear when uttering hating and liking, whereas male teenagers were more creative in their topics.

According to Rayson and Garside (2000), the process of Wmatrix approach follows these procedures. Given two raw corpora which are to be compared, a set of frequency lists for each corpus is generated. Previously, this would merely be a word frequency list, however, Rayson and Garside (2000) have established two lists, namely, a semantic tag and a part-of-speech frequency list. Owing to independence assumptions,

it is underlined that the two corpora must neither overlap between themselves nor be a sub-corpus of the other (Rayson, 2003: 95-96). Regarding every single word in the two frequency lists they have calculated the "log-likelihood (henceforth LL) statistic". The calculation is presented by construction a contingency table as in Table 3.1.

Table 3.1 Contingency Table for Log-likelihood Calculation

	Corpus One	**Corpus Two**	**Total**
Frequency of word	a	b	a+b
Frequency of word not occurring	c-a	d-b	c+d-a-b
Total	c	d	c+d

(Source: Rayson, 2003: 96)

The two formulas used to calculate the LL value are explained thoroughly using Rayson's (2003: 96-97) work afterwards:

> Note that the value 'c' corresponds to the number of words in corpus one, and 'd' corresponds to the number of words in corpus two (N values in the formula below). The values 'a' and 'b' are called the observed values (O). We need to calculate the expected values (E) according to the following formula:
>
> $$E_i = \frac{N_i \sum_i O_i}{\sum_i N_i}$$
>
> In our case N1 = c, and N2 = d. So, for this word, E1 = c × (a+b) / (c+d) and E2 = d × (a+b) / (c+d). The calculation for the expected values takes account of the size of the two corpora, so we do not need to normalise the figures before applying the formula. We can then calculate the log-likelihood value according to this formula:

$$-2\ln\lambda = 2\sum_i O_i \ln\left(\frac{O_i}{E_i}\right)$$

This equates to calculating LL as follows: LL = 2 ×((a ×ln (a/E1)) + (b ×ln (b/E2))).

The word frequency list is then sorted by the resulting LL values and the particular category in USAS standing at the top of the list indicating the words within that domain has "the most significant relative frequency difference between the two corpora" (Rayson, 2003: 97). The higher the LL value, the more significant gender differences are revealed. In this way, the results can manifest the extent of how much some words in one category are indicative (or distinctive) of one corpus, as compared to the other corpus from the place of log-likelihood value in the list.

Likewise, the next phase in the Wmatrix approach is to implement the same comparison at the POS and semantic levels. These comparisons broaden Scott's technique of WordSmith during 1996 to 1999 to produce key items rather than key words, and key grammatical categories and key concepts are also compared. At this point, the researcher must intervene and examine concordance examples of the significant words, POS and semantic tags highlighted by this technique. In this study, the researcher mainly analyses the comparison under key concept list and key POS list between the two genders.

The Wmatrix tags the items in three levels, namely, key words, key semantic domains and key parts of speech. “Semantic preference is the relation, not between individual words, but between a lemma or word-form and a set of semantically related words, and often it is not difficult to find a semantic label for the set” (Stubbs, 2001: 65). That is to say, the semantic level involves more context rather than merely focus on a particular word. As for part of speech, it supplements the occurrences of significant words and integrates content analysis.

3.3.2 Algorithm of Wmatrix Taggers

UCREL Semantic Analysis System (USAS)

Being the theoretical basis of Wmatrix, the semantic tags in USAS “show semantic fields which group together word senses that are related by virtue of their being connected at some level of generality with the same mental concept” (Archer, Wilson, and Rayson, 2002: 1). Those groups consist of antonyms, synonyms, hyponyms, as well as hypernyms. At present, the lexicon includes approximately thirty-seven thousand words and the template list includes at least sixteen thousand multi-word units. According to Archer et al. (2002), the semantic tags comprise the six following parts:

1. an upper case letter indicating general discourse field.
2. a digit indicating a first subdivision of the field.
3. (optionally) a decimal point followed by a further digit to indicate a finer subdivision.
4. (optionally) one or more ‘pluses’ or ‘minuses’ to indicate a positive or negative position on a semantic scale.

5. (optionally) a slash followed by a second tag to indicate clear double membership of categories.
6. (optionally) a left square bracket followed by 'i' to indicate a semantic template (multi-word unit) (p. 1).

The initial tagset, based on Tom McArthur's Longman Lexicon of Contemporary English (McArthur, 1981), provides the most suitable thesaurus type classification of word senses for this type of analysis. The latest tagset is permuted in a hierarchy with 21 major discourse fields broadening into 232 category labels (Archer et al., 2002). The 21 top levels of fields utilized by USAS are shown in the following table (Table 3.2).

Table 3.2 Semantic Domains Following the UCREL Semantic Analysis System (USAS)

A General and abstract terms	**B** The body and the individual	**C** arts and crafts	**E** emotion
F food and farming	**G** government and public	**H** architecture, housing and the home	**I** money and commerce in industry
K entertainment, sports and games	**L** life and living things	**M** movement, location, travel and transport	**N** numbers and measurement
O substances, materials, objects and equipment	**P** Education	**Q** language and communication	**S** social actions, states and processes
T Time	**W** world and environment	**X** psychological actions, states and processes	**Y** science and technology
Z names and grammar			

(Source: http://ucrel.lancs.ac.uk/usas/)

Constituent Likelihood Automatic Word-tagging System (CLAWS)

Currently, the majority of part-of-speech taggers are stochastic or probabilistic (Marshall, 1983; Garside, Leech and Sampson, 1987). Conventionally, they select a preferred tag for a word by "calculating the most probable tag in the context of the word and its immediate neighbours" (Garside and Smith, 1997: 102).

The CLAWS tagger, however, could be considered to be a hybrid tagger, involving both probabilistic and rule-based elements, even in its earliest form (CLAWS1-Marshall, 1983, Garside et al., 1987). The development of CLAWS1 started in 1980 at Lancaster University with an updated version of the Brown tagset[5], using about 135 tags (Garside and Smith, 1997).

The latest version of CLAWS is the CLAWS4 with a more scientific tagset used to tag the 100,000,000-word British National Corpus (BNC). The tagset contains two sub-tagset as Garside and Smith (1997: 108) state:

1) a detailed tagset (C7) of 146 tags for a two million word sampler corpus, and
2) a less refined tagset (C5) of 61 tags for the rest of the corpus.

[5] The original Brown corpus was the first million-word electronic corpus of English, created in 1961 at Brown University. This corpus consists of text from 500 sources, and the sources have been categorized by genre, such as news, editorial, and so on. Francis and Kucera (1979) established 75 non-compounded tags as the POS tagset for the corpus.

The present study establishes its theoretical basis on the USAS and CLAWS4 by using its SEMTAG and POS tag via Wmatrix3 which will be introduced in the next section.

3.4 Research Methodology

Through conducting a quantitative corpus-driven research on the content of Malaysian female and male blogs written in English, this project aims to answer the research questions raised. In this research, identical methodology is used to answer Research Question One and Two. The data of the corpus is acquired through the assistance of a software tool, namely, Wmatrix. Word ranking, word frequency, word percentage and word usage in both genders are compared.

3.4.1 The Corpus

The corpus of the research is coalesced by two sub-corpora. One is the sub-corpus of the female weblogs and the other one is the sub-corpus of the male weblogs. Complying with the standard data choice of Wmatrix (Rayson, 2003: 95-96), the two sub-corpora are not either overlapped or attached to the other one. Eventually, one hundred female weblogs (41, 585 words) and one hundred male weblogs (39, 222 words) are chosen as the target data and all of them are written in English by Malaysian locals.

Selection of Weblogs

In this research, the researcher examines one particular type of weblogs - personal blogs in Malaysia from http://www.blogmalaysia.com/ which is recommended by Amir et al. (2012). According to them, "BlogMalaysia was chosen because it is free and is the fastest growing blog directory in Malaysia" (Amir et al., 2012). The choice of this particular type is based on the fact that this category of weblogs is the most dominating in popularity (5292 logs) within this website which manifests that it is the most representative type to be analysed for this study. Moreover, personal blogs enjoy no boundary for specific topics. They are just the "daily diary" recorded by bloggers about the things they are interested in.

Presently, many social network services (SNS) such as Twitter, MySpace and Facebook, have become the craze among Netizens worldwide. With respect to the utilization of this social media within the ESL environment, the weblog has become progressively popular as a constructive and authentic learning tool especially in the language classrooms (Seitzinger, 2006). Furthermore, studies by Malaysian scholars indicate that weblog is effective in improving Malaysian ESL learners' language learning and writing skills (Bakar, Latif, and Ya'acob, 2010; Yunus et al., 2013).

According to Bakar et al. (2010), the use of blogs as classroom application provides avenue for the novices to be more independent and autonomous in mastering

the learning experience. Once they are acquainted with the gist of the advantages brought by weblogs, ESL learners can practice this skill beyond the classroom gradually. When students are able to engage with the target language more autonomously and frequently, blogs will offer them additional opportunities to acquire the language spontaneously.

For the corpus, the researcher establishes the criteria stated as follows:

1) the weblogs should be posted in the period of September, 2012 to April, 2013.
2) the minimum number of words for each blog should be two hundred.
3) the weblogs should be written in English or few Malay expressions could be acceptable.
4) the bloggers should post as least fifty weblogs in the records and start writing blogs on this website since two years ago.

By launching a systematic approach which will be illustrated in Section 3.4.2, two hundred personal blogs (80,807 word corpus) have been chosen to form the corpus of personal weblogs via this website in the period of September, 2012 to April, 2013. The sub-corpus for female weblogs contains 41,585 and the sub-corpus for male weblogs is 21,892 words with the minimum two hundred words in each selected blogs[6].

[6] Mulac and Lundell (1994), Nowson and Oberlander (2006), and Skoglund (2009) illustrate that females tend to produce longer texts than the male counterparts.

Profile of Bloggers

An identical number of male and female bloggers (one hundred each) constitute the participants of the research. On account of seeking weblogs in an ESL context, only blogs written in English by Malaysian adults are shortlisted for analysis. As for the author profiling, Santosh, Bansal, Shekhar, and Varma (2013) have proposed a "Machine Learning approach[7]" in order to identify the unknown author's age and gender. Based on the content, style and topic, the researcher has presented the updated profiles as follows:

Table 3.3 Profile of Bloggers

Gender	**20-30 years old**	**30-40 years old**	**40-50 years old**	**50 years old or above**
Female	39	26	14	6
Male	28	42	17	3

Moreover, all the selected bloggers have years of experience of blogging as the females started from the year 2003 to 2009 and males began from the year 2004 to 2008. Furthermore, their blogs have contained at least fifty postings in order to show that they are active bloggers.

[7] It is the approach to analyse the bloggers' identity based on the content, style and topic (Santosh et al., 2013).

3.4.2 Data Collection Procedures

To begin with, the researcher has browsed blogs on this website: http://www.blogmalaysia.com/. Of all the twenty-seven categories, *Personal* is the most popular topic within this website. The sorting of the topic is done in two ways. One is listed in an alphabetical order of each blog's title and the other one is sorted by PageRank (PR) created by Google.

A site's position within results returned for any search is very important, as many people rely only on the results on the first pages. Therefore search engines act as gatekeepers for certain information (Fallows, 2005: 36). As reported by Webcertain[8], *Google*, *Baidu*, *Yahoo*, *Yandex*, and *Bing* are the Top 5 leaders in worldwide search market share through 2012 which indicated that Google is used by the largest population worldwide. According to *2013 Search Engine Market Share By Country*, in Malaysia, Google's market share was about 93% in 2013[9]. PageRank has already been a "trademark" of Google as it is noted in Wikipedia:

> PageRank is a link analysis algorithm, named after Larry Page and used by the Google web search engine that assigns a numerical weighting to each element of a hyperlinked set of documents, such as the World Wide Web, with the purpose of "measuring" its relative importance within the set.
>
> (Source: http://en.wikipedia.org/wiki/PageRank)

[8] http://blog.webcertain.com/a-look-into-the-global-desktop-performance-of-the-5-biggest-search-engines-worldwide/12/02/2013/

[9] http://returnonnow.com/internet-marketing-resources/2013-search-engine-market-share-by-country/

As *Google* dominates the market of search engines, it can be concluded that PageRank is very important for the impact of any site in the Web. Pages with a higher PageRank tend to have higher visitor numbers, and it can be assumed that they therefore have more impact on one's attention (Kirchhoff, Bruns, and Nicolai, 2007). Therefore, the researcher is determined to use this sortation to collect her data.

In aid of this effective searching approach, 4687 of the 5292 personal logs were labelled by PR from 1 to 5. In order to systematically build the corpus, the researcher establishes three stages for selection process. At Stage I, the researcher scrutinized all the 697 records with a PR from 5 to 2, leaving out all the PR 1 logs. The reason for adopting this criterion was that the staple PR of the elementary data was PR 1 which made it impossible to judge the rationality of screening. At Stage II, the researcher handpicked the 426 blogs in the period of September, 2012 to April, 2013 to be in conformity with the norm of timeliness. At Stage III, only blogs written by Malaysian adults in pure English were selected due to the fact that English is the sole discernible language in Wmatrix. So far, 301 postings were verified as the target weblogs. Those blogs were randomly sorted in the posting list without division of the gender or time.

Ultimately, the researcher went through these target weblogs one by one until she verified the 100th female weblog and the 100th male weblog in the posting list. After narrowing down the two hundred sample blogs systematically, the researcher equally

divided them into two gender-oriented groups, namely, female sub-corpus and male sub-corpus, in order to ensure the precision and accuracy of the data.

3.4.3 Data Analysis Procedures

After sampling the data systematically, the corpus in two separate files was analysed by Wmatrix to obtain the underlying findings. Next, the researcher analysed the results by combing the relevant semantic domain into a more general topic and the relevant POS into a more general POS (such as noun, verb, adjective etc.).

Considering the Research Questions

In order to address Research Question One and Two, which were concerned with highlighting the main semantic and stylistic features between females and males in the two corpora, the researcher needed to decide on a number of analytical procedures. As mentioned earlier in Chapter 2 (Section 2.5), a distinction is often made between corpus-based and corpus-driven analyses (Tognini-Boneli, 2001: 65; 84). Within the corpus-based approach, researchers start with a theory, and then use the corpus to investigate the evidence in the data and testify whether it accords with their theoretic framework. In the corpus-driven approach, the researcher concern more about the data and its integrity without any hypothesis (Jawhar, 2012: 55).

As described earlier, it was decided that the latter type of analysis would be more

useful, as it would elicit differences and similarities that the research may not have considered in advance. The researcher therefore decided to use a number of automatic measures to ascertain important differences and similarities. First, key semantic domains were considered. Sometimes frequency differences between individual words could be too small for them to be key words. However, set of words which had similar meanings might contribute towards a larger difference when considered collectively (Baker, 2004).

Rayson (2008) and Culpeper (2009) also argued for the use of semantic domains in addition to key word analysis. This approach was then supplemented with an analysis of key parts of speech (POS). Using Wmatrix, the researcher was able to assign semantic and POS tags to her data to compare the two files in order to obtain significant USAS and POS tags.

Obtaining Lists

For the semantic domain analysis, texts were then analysed using the Wmatrix Semantic Tagger. Wmatrix is a web-based tool for corpus analysis and comparison of corpora by means of frequency lists, concordances, key grammatical categories and key semantic domains (Kaur, 2009: 103). Corpus texts are uploaded via a web browser where texts are automatically tagged for parts of speech using the CLAWS grammatical tagger, and the semantic domains using the USAS semantic tagger. The tagger compares

the frequencies of words in different semantic domains and parts of speech on log-likelihood (LL) tests. Two corpus files were uploaded onto the Wmatrix Tag Wizard creating two folders, namely, Malaysian female weblogs and Malaysian male weblogs.

Mistags

Wmatric assigns semantic tags to words by consulting an in-built lexicon. However, sometimes words can be mistagged, which could result in certain semantic categories appearing as key when they should not be (ibid: 103). To safeguard against this, the researcher manually examined all of the words in the two data sets in Wmatrix for any mistags. From the words in the domain, concordance lines were further observed to examine the context of the words. The purpose was to remove mistags from the semantic domain in order to ensure that only correctly tagged words remained. For mistags that were ambiguous, tags were checked with her supervisor. A list of mistags identified in the Malaysian corpora is shown in Appendix 3. A decision was made to re-calculate the log-likelihood scores of the semantic domains after removing the mistags (Section 4.2 shows the new semantic domains after recalculation of mistags).

The following figure illustrates the flowchart for the corpus analysis procedure from the preparation of the scripts in .txt files to the interpretation of the corpus findings.

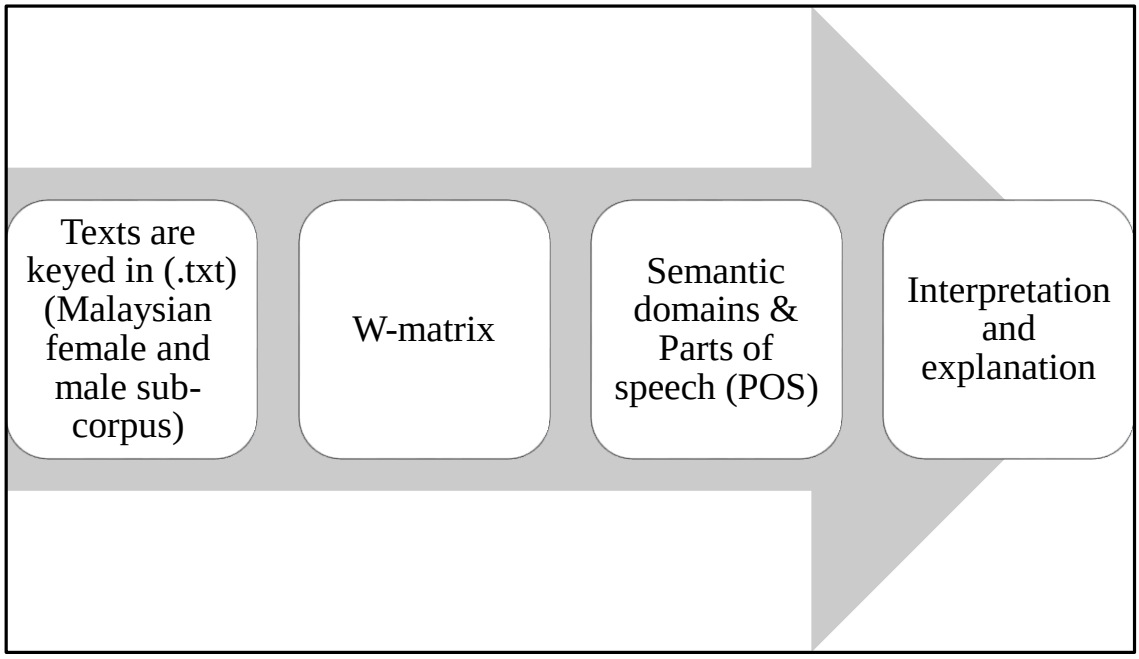

Figure 3.1 Flowchart of Data Analysis Procedures

The researcher further grouped the words from semantic domain into semantically-loose categories in order to examine them together (Section 4.3). As an auxiliary, the category of the key POS was classified into the basic types such as nouns, verbs, pronouns and prepositions (Section 4.4). The last stage of analysis, which is the qualitative part of the study, involves discussions of the results (Section 4.5).

Remarks on the Analysing Procedures

In this research, the data was analysed in the content of semantic domains tagged by Rayson's (2013) Semantic domains following the UCREL Semantic Analysis System (USAS) and the parts. The two sub-corpora of female and male personal blogs were then compared against each other using Wmatrix. The content of both men and women's personal weblogs were examined through comparing the frequency list,

following the highest value of LL (Log-likelihood) in Wmatrix. Furthermore, the word ranking, word frequency, word percentage and words usage in both genders are compared. Likewise, the two sub-corpora were compared in two phases; in Phase 1, female personal weblogs sub-corpus was compared to male personal weblogs sub-corpus whereas in Phase 2, male personal weblogs corpus was compared to female personal weblogs sub-corpus.

Wmatrix's website was accessed by logging into http://ucrel.lancs.ac.uk/wmatrix3.html. Both the female personal weblogs sub-corpus and male personal weblogs sub-corpus were transcribed into plain text and combined into two gender-based folders. Then the two folders were uploaded respectively into the software system.

Phase One began with the female personal weblogs sub-corpus being compared to the male personal weblogs sub-corpus and a key semantic domain list was obtained. Key semantic domains which had a LL value more than 10.83 were subsequently analysed. The higher the value of LL, the more significant differences were notified, when the female personal weblogs sub-corpus was compared with the male sub-corpus. The key domains and words, which were found to be interesting, would be analysed and argued in data analysis. Besides comparing word ranking, word frequency and word percentage, concordances would also be referred to in order to examine the usage of the

words in both genders. In Phase Two, sub-corpus of male personal weblogs was compared to female personal weblogs corpus while the rest of the methodology is similar to Phase One. The same procedure was applied in the analysis of key parts of speech (tagged by CLAWS).

To qualify for inclusion in the analysis, keyness for WordSmith 5.0 was set at $p<0.001$, and the log-likelihood critical value was applied at 10.83 for Wmatrix (Kaur, 2009). As the LL value list was sorted in descending order, the value above 10.83 ensure that the p value was less than 0.001 (Rayson, Berridge, and Francis, 2004). A further qualitative analysis was then made to explain the findings with 'Difference Theory' and 'Dominance Theory' (Nemati and Bayer, 2007: 136-137).

Nemati and Bayer (ibid: 136) illustrated that that in "difference theory", although men and women live in the same environment they establish different relations with the society as if each belongs to a different environment and culture, the result of which is consequently reflected in the language of both genders. This is the cultural factor that causes gender differences. In "dominance theory," men and women are believed to inhabit a cultural and linguistic world, where power and status are unequally distributed. This is the social factor that produces gender differences (ibid: 137). The methodological framework of the present study is shown in Figure 3.2.

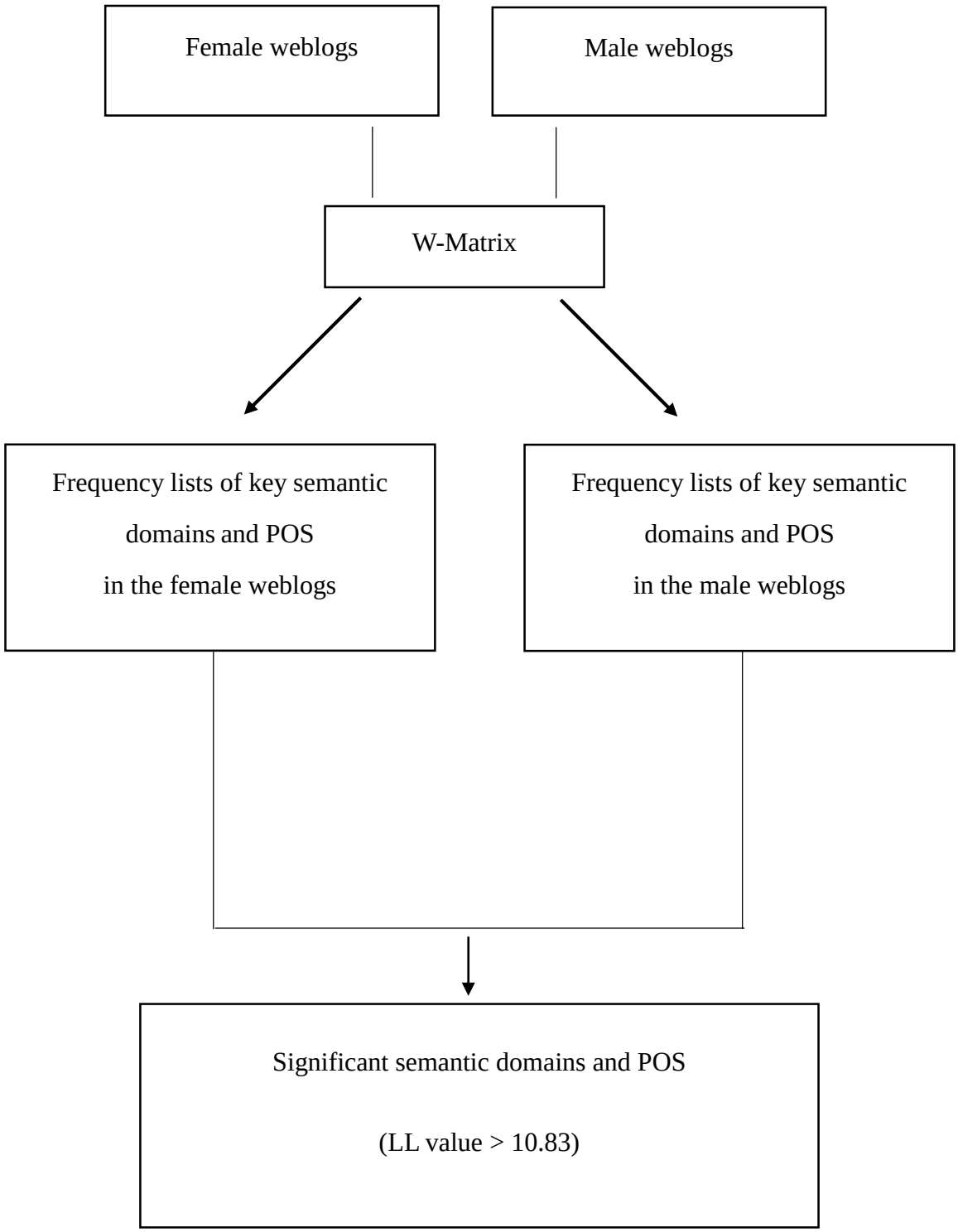

Figure 3.2 Methodological Framework of the Present Study

CHAPTER 4

ANALYSIS AND DISCUSSIONS

4.1 Introduction

In this analysis chapter, the corpus of weblogs are analysed using the corpus-driven approach as conceptualised in the previous literature review chapter (Section 2.5). This chapter is divided into five sections. The analytical framework is established in Section 4.2. Then, Section 4.3 presents the analysis of the key semantic domains along with key parts of speech in Section 4.4. Finally, a qualitative analysis of the findings is elaborated in Section 4.5.

4.2 Analytical Framework

As mentioned in the previous chapter, the 80,807-word corpus built for this research was based on two hundred English weblogs of Malaysian adults which were collected from the website http://www.blogmalaysia.com/. In order to answer Research Question One and Two, the parameters of two corpus analysis were incorporated, namely a detailed examination of key semantic domains and key part of speech, both of which also took into consideration concordance lines, collocates, and clusters of the word under analysis. These analyses were based on the application of the corpus tool: Wmatrix as shown in Figure 4.1.

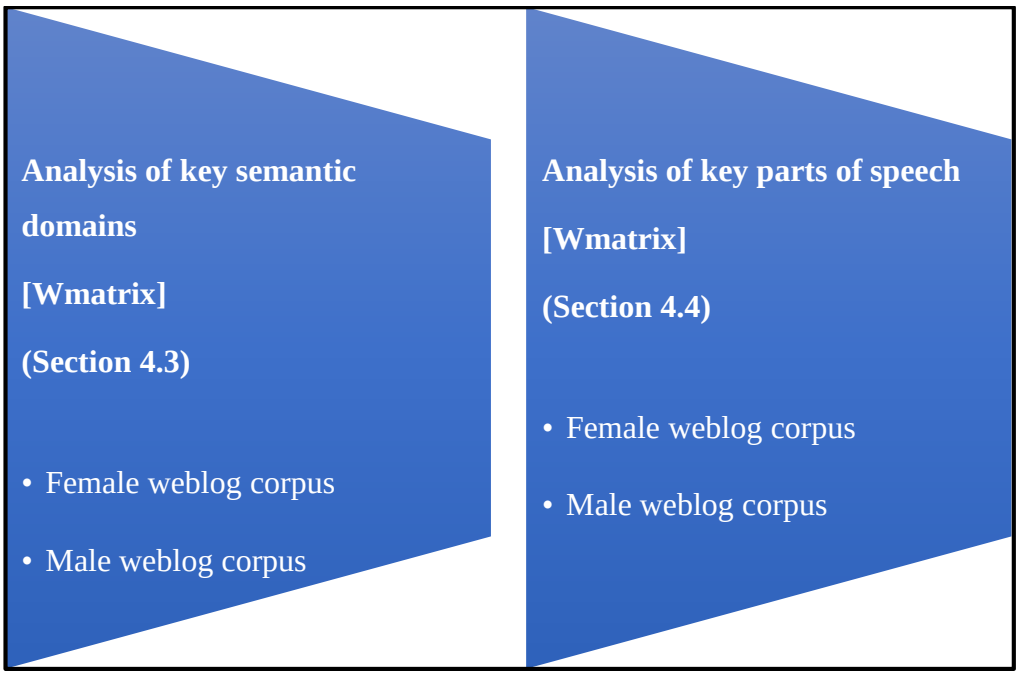

Figure 4.1 Analytical Framework of the Data Analysis

Due to space limitations, if a word's frequency was above ten, then the researcher would normally only present a representative sample of ten concordance lines in this chapter (although in some cases she presents more than ten lines, especially for very frequent words). However, if there were fewer than ten concordance lines, the researcher includes, where possible, each line in her analysis. Following the same methodology, four sub-questions of Research Question One and Two are answered in Section 4.3 and Section 4.4.

4.3 Analysis of Key Semantic Domains

This section presents the analysis of the key semantic domains in the Malaysian female and male sub-corpus. As an effective tool, Wmatrix was used to assign semantic

tags automatically to the female and male sub-corpus to compare significant semantic domains in one sub-corpus with its reference sub-corpus. The same cut-off log-likelihood value as for the semantic comparison was set at 10.83 to compare the female sub-corpus with the male sub-corpus. The researcher first discussed the key semantic domain in the female sub-corpus (Section 4.3.1) followed by the key domains significant in the male sub-corpus (Section 4.3.2).

4.3.1 Key Semantic Domains in the Female Weblogs Sub-corpus

Given the female sub-corpus (see Chapter 3 for a further description of the corpora), Wmatrix tagged 41,585 words (95.66%) from the original texts, or 43,471 words. In total, the tagger generated 369 semantic categories in total. The researcher examined all the concordance lines for the words categories in each of the semantic domains of the female sub-corpora to identify mistagged items, and removed the mistags from the analysis as they were not relevant based on the context of the word (see Appendix 3 for a list of mistags in the female sub-corpus). The distributions of the twenty-one top level of fields are shown in the following table which is sorted by frequency to show key items at the top (Table 4.1). The fields of the same frequency are listed in alphabetical order.

Table 4.1 Frequency Lists for Fields of Semantic Domains in Female Weblogs

Rank	Fields of Semantic domains	Frequency	%
1	**A**[10]: General and abstract terms	80	21.68
2	**N**: numbers and measurement	45	12.20
3	**S**: social actions, states and processes	44	11.92
4	**X**: psychological actions, states and processes	42	11.38
5	**T**: Time	25	6.78
6	**I**: money and commerce in industry	19	5.15
7	**E**: emotion	17	4.60
8	**O**: substances, materials, objects and equipment	17	4.60
9	**G**: government and public	10	2.71
10	**Q**: language and communication	10	2.71
11	**B**: The body and the individual	9	2.44
12	**Z**: names and grammar	9	2.44
13	**K**: entertainment, sports and games	8	2.17
14	**M**: movement, location, travel and transport	8	2.17
15	**H**: architecture, housing and the home	6	1.63
16	**W**: world and environment	6	1.63
17	**F**: food and farming	5	1.36
18	**L**: life and living things	5	1.36
19	**Y**: science and technology	2	0.54
20	**C**: arts and crafts	1	0.27
21	**P**: Education	1	0.27

In this section, the female sub-corpus is analysed to examine the key semantic domains that are significant in the male sub-corpus in comparison with the female sub-corpus. Concerning the log-likelihood value, three semantic fields, namely, *M: movement, location, travel and transport*; *P: Education*; and *Y: science and technology*, are not tagged which indicates that there is no gender difference in those semantic categories when the female sub-corpus are in relation to the male sub-corpus.

[10] These capital letters refer to the USAS tagger in Table 3.2.

Table 4.2 Key Semantic Domains in Female Weblogs Compared to Male Weblogs

Rank	**Tag**	**Semantic domain**	**Female weblogs**		**Male weblogs**		**LL value**
			Freq	**%**	**Freq**	**%**	
1	F1	Food	636	1.53	343	0.87	72.75
2	Z8	Pronouns	5567	13.39	4529	11.55	54.82
3	B3	Medicines and medical treatment	165	0.40	65	0.17	39.31
4	S4	Kin	282	0.68	145	0.37	37.09
5	S2.1	People: Female	67	0.16	14	0.04	34.68
6	B1	Anatomy and physiology	363	0.87	224	0.57	25.60
7	B2-	Disease	137	0.33	62	0.16	24.76
8	H2	Parts of buildings	82	0.20	28	0.07	24.63
9	T3--	Time: New and young	38	0.09	9	0.02	17.59
10	C1	Arts and crafts	153	0.37	83	0.21	17.19
11	L2	Living creatures: animals, birds, etc.	137	0.33	72	0.18	16.93
12	O4.4	Shape	66	0.16	25	0.06	16.83
13	A9+	Getting and possession	660	1.59	501	1.28	13.54

As shown in Table 4.2, Wmatrix revealed thirteen semantic domains (LL value > 10.83) that emerged from the analysis in which the adjusted frequencies and log-likelihood value are calculated. Although the fields of *A: General and abstract terms*; S*: social actions, states and processes* and *X: psychological actions, states and processes,* dominate the frequency list of the semantic fields, they are not of significant gender differences when the females are compared to the males as identified by the log-likelihood value. The field of *N: numbers and measurement* is not even displayed in Table 4.2 which indicates that the difference is not so salient. The researcher conducts her analysis by taking the female sub-corpus as a cornerstone. As an auxiliary, the male semantic domains likewise are discussed where necessary.

Table 4.2 reveals an interesting relationship between the semantic domains with the strongest key semantic domain being *Food (F1)* and the log-likelihood value is 72.75. There are altogether ten semantic categories in the table, including *F: food and farming (F1)*; *Z: names and grammar (Z8)*; *B: The body and the individual (B3, B1, B2-)*; *S: social actions, states and processes (S4, S2.1)*; *H: architecture, housing and the home (H2)*; *T: Time (T3--)*; *C: arts and crafts (C1)*; *L: life and living things (L2)*; *O: substances, materials, objects and equipment (O4.4)* and *A: General and abstract terms (A9+)*. The most frequent number of items found in a semantic domain is *Pronouns (Z8)* which is also the second entry in Table 4.2 and the LL value is 54.82. The males (4,529 words, 11.55%) seem to use pronouns less frequently than females (5,567 words, 13.39%) although in terms of percentage difference there is only a 1% difference. Malaysian females also tend to make references to human body and healthy issues, including *Medicines and medical treatment (B3)*, *Anatomy and physiology (B1)* and *Disease (B2-)*, references to people and relationships, such as *Kin (S4)* and T*ime: New and young (T3--)*, references to *Parts of buildings (H2)*, *Arts and crafts (C1)*, *Living creatures: animals, birds, etc. (L2)* and *Shape (O4.4)*.

In order to fulfil the task of analysing the twelve semantic domains more manageably, by taking Schmid (2003) and Kaur (2009)'s classification, the researcher further grouped them based on themes that tend to relate to each other, which resulted in four main groups: *People and relationships*; *Body and health*; *Food* and *Others* (see

Appendix 4 for the ten most frequent terms in each domain). In the following parts, the researcher discussed mainly the language used in female sub-corpus, although references were also made to male sub-corpus to highlight any interesting observations, where necessary.

People and Relationships

There are four semantic domains that fall under this category, namely, *Pronouns* (Z8, LL value = 54.82); *Kin* (S4, LL value = 37.09); *People: Female* (S2.1, LL value = 34.68) and *Time: New and young* (T3--, LL value = 17.59). The researcher has grouped the T3- domain under this category for the reason being that at first glance, it seems that the theme of this domain concerns time. However, upon closer examination of the types, and the concordance lines in which they occur, the words are associated with people. Words under the T3-- tag such as *baby* (32 times, 0.08%) and *younger* (4 times, 0.01%) related to the female focus on family in their writing.

The females use pronouns in the first person most frequently as demonstrated by the two most frequent pronouns in the Z8 tag: *I* (1392 times, 3.35%) and *my* (586 times, 1.41%). The males use the first person pronouns less frequently than the females: *I* (1133 times, 2.89%) and *my* (418 times, 1.07%). There is also a large proportion of *we* as tagged in the female (283 times, 0.68%) and male sub-corpus (236 times, 0.60%). As for the gender-featured words, the use of *he* (94 times, 0.24%) is more frequent

compared to *she* (19 times, 0.05%) in the male sub-corpus but the former is used to a greater extent in the female sub-corpus (166 times, 0.40%). This seems to differ with the results of Argamon et al. (2003) that male authors use more male third-person pronouns in writing. On the contrary, in the female sub-corpus, the use of *she* (121 times, 0.29%) is slightly less frequent than *he* (166 times, 0.40%) but is still considerably more than that of the male (19 times, 0.05%).

Linked to the high use of *she* by females is the fact that they make more reference to related females as shown in the *People: Female (S2.1)*. This is marked by a high frequency of *girl* (17 times, 0.04%), *girls* (14 times, 0.03%) and *lady* (13 times, 0.03%). In the male sub-corpus, there are infrequent reference to *girl* (5 times, 0.01%), *girls* (once) and *lady* (once). Furthermore, the top ten most frequent type list of this domain, i.e. *People: Female (S2.1)* manifests that the females use more pronouns in the first person and the third person than the males, but there are fewer pronouns in the second person tagged in the female sub-corpus than in the male sub-corpus. There is only one lexical word in the second person, namely, *you* (310 times, 0.75%) on the female list whereas *you* (349 times, 0.89%) and *your* (123 times, 0.31%) on the male list.

The females also refer to more kinship terms than the males as is evident from the *Kin (S4)* semantic domain which is higher in the female blogs (0.68%) relative to the male blogs (0.37%). Nowson (2006) notified that human and family are adored by the

females and Newman et al. (2008) also clarified that females make references to home and family. There are much higher proportions of references to mums with the fourth most frequent terms of address being *mom* (17 times, 0.04%), less frequent terms include *mama* (8 times, 0.02%), *mother* (6 times, 0.01%), *mum* (6 times, 0.01%), *mommy* (5 times, 0.01%), *mummy* (5 times, 0.01%), and the plural, *moms* (once). Not surprisingly, the females also use more female kinship words including *sister* (16 times, 0.04%), *grandma* (8 times, 0.02%), *niece* (4 times, 0.01%) and *aunt* (twice). However, the females also refer to male relatives such as *brother* (21 times, 0.05%), *father* (21 times, 0.05%), *dad* (8 times, 0.02%), *nephew* (6 times, 0.01%) and *grandpa* (twice). On the other hand, the males refer to female relatives less frequently in their blogs, i.e. *aunt* (5 times, 0.01%), *mum* (5 times, 0.01%), *sister* (3 times, 0.01%) and *granddaughter* (twice, 0.01%).

Additionally, the females also use colloquial form of kinship terms such as *hubby* (12 times, 0.03%), *sis* (twice), *bro* (once), *daddy* (once) and *auntie* (once) to show intimate relationship.

Referring to the fact that the research focuses on adults' weblogs, marital relationship serves as a critical segment in the kinship's domains. The females seem to

be more concerned about their spouses by using *husband* (14 times[11], 0.03%) and *hubby* (12 times, 0.03%) and less about themselves as the related word, *wife*, is only used 6 times (0.01%). This predilection is also shared by the males who use the word, *wife* (12 times, 0.03%), much more than the word *husband* (twice, 0.01%). Nevertheless, the males also refer to marriage by using words including *marriage* (8 times, 0.02) and *wedding* (6 times, 0.02) which are not traced in the female key words. As for the parent-child relationship, the females mention more about their sons than daughters by using the word '*son*' (10 times) and the word '*daughter*' (4 times) which share equal frequency (7 times) in the male sub-corpus.

Concordance 4.1 Concordance lines (females) of *nephew* (all 6 lines)

N			
1	pump! Dear cousin and dear	*nephew*	were delighted to see this! Haha!
2	to the Australian Garden !" Dear	*Nephew*	dear Nephew, wanna go there
3	Garden!" Dear Nephew ah, dear	*Nephew*	, wanna go there ah?? Since we a
4	Huh? Huh? Huh?" Dear	*Nephew*	looked at me and then turned in !
5	children last week. My niece and	*nephew*	showed me their painting, on
6	movies also. My husband's	*nephew*	and niece also surf there, one

Kinship is portrayed in relation to the roles of different family members in the lives of the females themselves and how the females are affected by the presence of family members. For example, *nephew* (6 times, 0.01%) figures someone who has a distant relationship, yet is connected to the writer through his behaviour or existence as shown

[11] There are eleven concordance lines talk about the blogger's own husband and the other three lines refer to others' husbands.

in all six examples of concordance line of the word *nephew* from the female sub-corpus (Concordance 4.1).

Concordance 4.2 Concordance lines (females) of *brother and sister* (20 of 37 lines)

N			
1	re eyes! So pretty right? But my	*brother*	said she's creepy. I'm heading to
2	asked for opinions from my	*brother*	and Chun. Discussed for the
3	with? My sister and my	*brother*	, my (only) eldest brother who ha
4	nd my brother, my (only) eldest	*brother*	who had just arrived from San
5	ng Ramli. The last time my eldest	*brother*	was here, we had dinner at the
6	in the itinerary. My sister, my	*brother*	and I arrived first and thus had
7	ati kurang pedas crabs. My eldest	*brother*	and Abang Ramli were running
8	they flew to United States for my	*brother*	's graduation. Happy New Year!
9	observant enough. And then my	*brother*	and I were trying to act to get a
10	frame for you to pose. P/S: My	*brother*	's girlfriend was really brave to
11	family members around. My	*sister*	and cousin came to visit
12	all too good about it. Even my	*sister*	is not answering my text. I
13	people we celebrated with? My	*sister*	and my brother, my (only) eldest
14	urian buffet in the itinerary. My	*sister*	, my brother and I arrived first an
15	was happening close to where my	*sister*	and I were sitting. He had to cons
16	. Boyfriend Ken pick me and my	*sister*	from Desa park; we went Jaya
17	Like a teenager again I visited my	*sister*	and her children last week. My
18	gle" days, goes shopping with my	*sister*	till we drop, try out restaurant,
19	finally decided to see my elder	*sister*	's dentist for consultation. Last
20	example of the fail attempt. My	*sister*	and I do not have the same hair

The females, however, more frequently refer to siblings, i.e. *brother* (21 times, 0.05%), *sister* (16 times, 0.04%), *cousin* (3 times, 0.01%) and *siblings* (twice) in describing their stories with their own siblings (Concordance 4.2). But again, those words are not used in the male sub-corpus of weblogs.

Concordance 4.3 Concordance lines (females) of *family, married, divorcee,* and *divorced* (5 of 40 lines)

N			
1	demise. I was too busy raising a	*family*	in Kuala Lumpur to keep track
2	busy career, helping to run a	*family*	. I never felt worrying about
3	By the way, announcing, I AM	*MARRIED*	yeah uuu. Syukur alhamdulilllah
4	the bus, I thought," Another	divorcee	, they are alone and not with
5	properties and she get hurt. We	*divorced*	and my three daughters stay

Further, and still on *Kin (S4)* domain, various influences in family relationships contribute to the female construction of the family unit which includes the relationship between *parents* (9 times, 0.02%) and children (*son*: 10 times, 0.02%; *daughter*: 4 times, 0.01%), and the relationship between *husband* (14 times, 0.03%) and *wife* (6 times, 0.01%). For example, in Concordance 4.3, if we take the most frequently-used word, *family* (31 times, 0.07%), socioeconomic status[12] seems to be a concern for the females as they describe how they raise the family as revealed in Lines 1 and 2. Gambell and Hunter (2000) noticed that the females tend to write more about romance. Peterson (2002) also revealed that romance and relationships are mainly dealt by the females.

The females in the present study use personal experiences and evidence from real life to consider married relationships not only with a fairy-tale ending of happily ever after, e.g. *wedding* (7 times, 0.02%), *married* (6 times, 0.01%), *weddings* (twice), *bridal* (once), *marry* (once), *engaged* (once), *getting married* (once) and *got married* (once),

[12] Bimber (2000) argued that socioeconomic status is the initial cause of gender difference.

but also divorce, e.g. *divorcee* (twice), *single parents* (once), and *divorced* (once). On the other hand, conflicts between husband and wife, or between parents tend to occur in descriptions of real situations and are upsetting experiences for the females.

There are only three types under *Time: New and young (T3--)* domain in the female sub-corpus, namely, *baby* (32 times, 0.08%), *younger* (4 times, 0.01%) and *toddler* (twice). These words, by contrast, do not relate to time, but instead to the new-born infants or the family members. Exceptionally, the word *younger* is once used in association with the generation, as one female writes 'If I have time I rather cook to have a good meal than wasting time surfing, as I belong to the older generation, not like the younger generation who carry their ipad 24 hours'. Coincidently, a male also uses the word *younger* in the same context: 'The restaurant was beautifully built and peppered with chic designs to attract the younger generations to enjoy their not-so-healthy meals there'. The males share the first three types (*younger*: 3 times, 0.01%; *baby*: 3 times, 0.01; and *toddler*: once) with the females but they also use *baby stuff* and *toddlers* once respectively of which are all referred to babies.

Body and Health

There are three semantic domains grouped under this category in the female sub-corpus, namely, *Medicines and medical treatment (B3)*, *Anatomy and physiology (B1)* and *Disease (B2-)*.

At position three, is the B3 tag which is used to a greater extent in the female sub-corpora compared to the males. This includes words referring to places: *hospital* (18 times, 0.04%), and *ward* (8 times, 0.02%); medicines: *balm* (13 times, 0.03%), and *capsules* (5 times, 0.01%); therapy: *treatment* (10 times, 0.02%), *hospitalized* (8 times, 0.02%), *surgery* (7 times, 0.02%) and *massage* (6 times, 0.01%); and staffs: *doctor* (8 times, 0.02%) and *dentist* (6 times, 0.01%). Arici (2009) indicates that the girls prefer to write about health than boys. Aside from that, the males seem to be more sensitive than the females during the therapeutic process. Taking the concordance lines of *dentist* as an example, the males express fear and anxiety unstintingly but there is only one reference to the uncomfortable feeling in the female sub-corpus but ends in a satisfying condition.

Text 4.1 Script Female 91

My first visit with Dr Lee was a surprisingly amazing one. I always dread going to the ***dentist***. It is a traumatic experience I try to avoid at all cost. I was very scared to have my tooth extracted - but know what? It was totally painless!

Dr Lee first put a surface anaesthetic gel on to my gum and after a few minutes, it was totally numb. Dr Lee injected my gum using the 'Magic Wand', without me feeling it, and skilfully extracted the tooth. The injection was painless.

I left the gauze in for about half an hour. I was told to gargle very gently every hour for the first day. With a numb right jaw, I could even eat *Bak Kut Teh* for lunch after the extraction. : D

By 3.18 pm, the numbness was gone. I didn't feel any pain. Amazing!

After 24 hours, I was fine. Wow, technology and science has come a long way the last couple of years.

Text 4.1 tells the story of how the female blogger sets aside her prejudice towards the dentist, yet the following concordance lines describe the male's suffering.

Concordance 4.4 Concordance lines (males) of *dentist* (5 of 10 lines)

N			
1	r my following blog post! Fear of	*dentist*	. It was quite a tough week
2	little fear in me of going to the	*dentist*	doesn't help. I was wondering
3	I was wondering would the	*dentist*	needed an operation, as this is a
4	, I was really that fearful of	*dentist*	and even bravely went for my
5	se are done, I began avoiding the	*dentist*	till last Friday. I was quite

As for the domain, namely, Anatomy and physiology (B1), gender differences lie in the attention to body parts. In the sub-corpus, both the females and males take account of sleep and hence the word *sleep* is frequently used in the two sub-corpora (15 times in the male sub-corpora and 12 times in the female sub-corpora). Furthermore, the females show greater interest in tiredness as they use the adjective *tired* 16 times where the males solely use it 5 times. The females concern more about their appearance, therefore, those beauty-related words consist of *eyes* (15 times, 0.04%), *hair* (12 times, 0.03%), *skin* (7 times, 0.02%) and *face* (6 times, 0.01%) emerge continually. In contrast, the males prefer to mention limbs with reference to words such as *hands* (10 times, 0.03%), *legs* (7 times, 0.02%) and *foot* (4 times, 0.01%).

With regards to pains in the individual's body, this study shows that the females blog more about suffering from neck pain, shoulder pain, heartache and backache while the males suffer more from headaches. Consequently, *neck* (19 times, 0.05%), *shoulder* (17 times, 0.04%), *heart* (11 times, 0.03%) and *back* (9 times, 0.02%) are unceasingly used in the female sub-corpus whereas *head* (8 times, 0.02%) is tagged in the male

sub-corpus. The males also use the word *tooth* (5 times, 0.01%) with all related to toothache as shown in Concordance 4.5.

Concordance 4.5 Concordance lines (males) of *tooth* (all 5 lines)

N			
1	d. Thus, I decided to 'tahan' the	*tooth*	for the entire week leading to my
2	will likely be an extraction as the	*tooth*	has chipped. A portion of the
3	ooth has chipped. A portion of the	*tooth*	actually came out. If it's an extr
4	to 'implant' back the missing	*tooth*	. After all these are done, I began
5	and also take time to explain my	*tooth*	's problem on a LCD screen and

Regarding the seventh most significant semantic domain, Disease (B2-), the females only name one specific type in the top 10 word list, namely, *asthma* (24 times, 0.06%), which is also the most frequently used word under this tag. In the female sub-corpus, females use words such as *pain* (14 times, 0.03%), *disabled* (6 times, 0.01%), *ill* (4 times, 0.01%), *painless* (4 times, 0.01%), and *madness* (3 times, 0.01%). However, males mention about *fever* (5 times, 0.01%), *dengue* (an epidemic disease in tropical areas, 4 times, 0.01%), *cancer* twice, 0.01%) and *paralysis* (twice, 0.01%). This indicates that males prefer stating their illnesses themselves, while females tend to describe their uncomfortable feelings.

Food

The semantic domain *Food (F1)* is the most significantly-used semantic category for females when compared to males (LL value 72.75). The females (636 words, 1.53%)

use words in this category to a greater extent than the males (343 words, 0.87%). This is largely demonstrated by the word that encompass the meals including *dinner* (21 times, 0.05%), *breakfast* (18 times, 0.04%), and *lunch* (15 times, 0.04%). They also relate to their cooking experience by using words such as *cook* (21 times, 0.05%), *recipe* (21 times, 0.05%) and *cooking* (16 times, 0.04%).

Owing to their responsibility to feed the family, the females focus on how to make the food rather than to eat it by using verbs including *fried* (10 times, 0.02%), *cooked* (8 times, 0.02%), *bake* (4 times, 0.01%) and *roasted* (3 times, 0.01%) as well as nouns related to cooking, such as *recipe* (21 times, 0.05%), *kitchen* (5 times, 0.01%), *skillet* (3 times, 0.01%), *crockery* (twice) and ingredients: *garlic* (9 times, 0.02%), *sauce* (8 times, 0.02%), *salt* (6 times, 0.01%), *sugar* (4 times, 0.01%), *mustard* (3 times, 0.01%) and *curry powder* (3 times, 0.01%).

While the females frequently write about how they cook, which is a stereotypical female behaviour (Quicke, 2012), the males, on the other hand, refer to what to eat, with *food* (36 times, 0.09%) as the most frequent word followed by *rice* (8 times, 0.02%), *meat* (8 times, 0.02%), beef (6 times, 0.02%), pork (6 times, 0.02%), seafood (5 times, 0.01%), soup (5 times, 0.01%) and BBQ (3 times, 0.01%). Being the most significantly used word in the female sub-corpus, *eat* (30 times) is closely linked with kids according to the female blogs (Concordance 4.6).

Concordance 4.6 Concordance lines (females) of *eat* (12 of 30 lines)

N			
1	or a while that he won't be able to	*eat*	breakfast at school when
2	us find it hard to teach our kinds to	*eat*	vegetables and fruits, we often give
3	yourself! If you want your kids to	*eat*	raw cucumber, tomatoes, carrots, s
4	toes, carrots, salad etc, you	*eat*	it first, let your kids watch you eat
5	t it first, let your kids watch you	*eat*	, and then offers it to them. If the
6	how those kids at Kindergarten	*eat*	raw vegetables, and they didn't jus
7	w vegetables, and they didn't just	*eat*	it, they enjoyed it! And at Afeef'
8	ehehe.. By encouraging your kids t	*eat*	raw vegetables, you'll get two bene
9	ic getting her dolls to sit down and	*eat*	.. kept on calling us " Mama! Sit do
10	on calling us " Mama! Sit down	*eat*	!" That night, she slept with her d
11	as told by the elderly, if kids can	*eat*	, let them eat. They know their
12	elderly, if kids can eat, let them	*eat*	. They know their tummy better.

The concordance lines above show that 12 of 30 occurrences are related to children's feeding issues. However, there is no occurrence that is relevant to children but is interrelated with rational diet of the blogger himself in the male sub-corpus (Concordance 4.7).

Concordance 4.7 Concordance line (males) of *eat* (all 12 lines)

N			
1	worry about what you are going to	*eat*	and drink. Do not worry about what
2	. Do not keep saying, What will we	*eat*	? or, What will we drink? or, What
3	m is I do a lot of sports but then I	*eat*	a lot. This is probably the last pos
4	to know the general idea of what to	*eat*	, how many calories of foods from
5	motivates me to keep working hard,	*eat*	clean and most importantly
6	of cabbage leaves, because we	*eat*	bulgogi. I heard that it is little s
7	Question which is: What did I	*eat*	for dinner on April 1st, 2013? Answ
8	with beef) as shown above is a mus	*eat*	! And so is the bun cha (grilled por
9	Maybe. And some of the locals	*eat*	it. If you are squeamish, you may n
10	re hung out there for all to see and	*eat*	besides the dog? Well, you might
11	, we have to pay our fees before yo	*eat*	the food. The lunch table was
12	r u a woman? Whr's ur abs? y u no	*eat*	protein? Y u still fat? ' knnccb.

Others

The researcher has grouped the other five semantic domains under this category as they do not relate to the other categories. They are *Parts of buildings (H2)*; *Arts and crafts (C1)*; *Living creatures: animals, birds, etc. (L2)*; *Shape (O4.4)*; and *Getting and possession (A9+).*

1. Parts of buildings (H2)

The significant gender difference lies in the use of the word *room*. Not only do they differ in the number of the word, (the females use the word *room* 20 times, the males use *room* 4 times), they also differ from each other in the description. A majority of the concordance lines (8 out of 20) show females refering room in relation to its functions in resting (Concordance 4.8).

Concordance 4.8 Concordance line (females) of *room* (8 of 20 lines)

N			
1	leep. We had mosquito coils in our	*room*	, actually I took the pillow from th
2	s to do, I rather spend my life at	*room*	Yeah it sounds so damn free and
3	To be a queen size bed in in Tok's	*room*	In which I would sleep alongside
4	That Sheraz has to sleep in his	*room*	. When he was 6 months old, we
5	sleeping for three weeks in his own	*room*	. We purposely do it now, not later
6	hubby accompanied both kids to the	*room*	and get them to sleep. Minutes late
7	for the week. Walking towards the	*room*	, I hear sounds. sounds that make
8	ge session and sleeping in the same	*room*	with a roomie who is pretty much

In contrast, in the male sub-corpus, *room* is used for daily routines and leisure activities as shown in Concordance 4.9.

Concordance 4.9 Concordance line (males) of *room* (all 4 lines)

N			
1	de us to be different and allow us	*room*	to individually grow at a pace
2	to own myself a PS3 and plat at my	*room*	to excel my prowess at Street
3	re bottled water in the car, in my	*room*	, under my desk. Minimal daily
4	have clothes scattered all over my	*room*	trying to decide what to wear for

Since the females link their rooms to sleep and rest, they focus on the domiciliary functions by using words like *department* (7 times, 0.02%), *bedroom* (4 times, 0.01), *suite* (twice) and *living room* (twice). As the males are busy with everyday concerns, they use words related to activities such as *reception* (3 times, 0.01%), *locker room* (once) and *ballroom* (once).

2. Arts and crafts (C1)

Overall, the females (153 types, 0.37%) use the words that depict arts and crafts twice as many as the males (83 types, 0.21%). Both the gender use the words *photo* (F: 26 times, 0.06%; M: 4 times, 0.01%), *photos* (F: 14 times, 0.03%; M: 8 times, 0.02%), *pictures* (F: 14 times, 0.03%; M: 5 times, 0.01%) and *picture* (F: 9 times, 0.02%; M: 6 times, 0.02%) most significantly. Another relevant word in the field of photography, namely, *camera* (10 times, 0.02%) is of high frequency accordingly. Moreover, the females use words like *art* (14 times, 0.03%), *painting* (7 times, 0.02%), *design* (5 times, 0.01%) and *designed* (5 times, 0.01%) which are also found in the male sub-corpus but with lesser frequency. On the other hand, the males use many culture-related words

including *cultural* (4 times, 0.01%), *cultures* (4 times, 0.01%), and *culture* (twice, 0.01%) which are untraceable in the female sub-corpus.

3. Living creatures: animals, birds, etc. (L2)

The word *chicken* is the most frequently used in both sub-corpora (female: 29 times, 0.07%; male: 8 times, 0.02 %). This is probably because chicken is the principal food of meat products in Malaysia (Jayaraman, Munira, Dababrata and Iranmanesh, 2013). Therefore, all the concordance lines of chicken are related to the food other than the animal itself.

As for wildlife, the males write more savage animals than the females by using the words *bear* (twice, 0.01%), and *snake* (twice, 0.01%) while the females merely use the word *tiger* (5 times, 0.01%) in the top ten list of this domain. To be more specific, four of the five concordance lines are linked to the name of a massage club which are considered mistags. Even the last one, 'The Tiger Show', mentioned in the text is NOT an animal show but a performance by nude dancers (Text 4.2).

Text 4.2 Script Female 95

Apart from intricately designed temples, beautiful blue ocean and the city beautiful of Bangkok, what else do you think of when you hear the word "Thailand"? Well, a bunch of the more adventurous people in this trip decided to check out ***the Tiger Show***. Though I don't think it's necessary for me to describe in detail the things that I saw in the show, let's just say, tiger is not the only "animal" available in this show.

Furthermore, the bloggers also recount edible animals and byproducts. The females use the words such as *crabs* (7 times, 0.02%), *eggs* (6 times, 0.01%), *crab* (5 times, 0.01%) and *fish* (5 times, 0.01%) where the males use *fish* (3 times, 0.01%), *mussels* (twice, 0.01%) and *cows* (twice, 0.01%). They also characterize some homebred animals such as *bunny* (4 times, 0.01%) in the female blogs and *turtle* (twice, 0.01%) in the male blogs. The gender differences in referring to animals may seem to be affected by the gender difference in human-animal interactions (Herzog, 2007). The slight differences between Herzong's (2007) result and the present one is that animal hunting is an evident activity in the Euro-American countries, whereas there is no relevant clues in the male sub-corpus in Malaysia.

4. Shape (O4.4)

In general, the words in the female sub-corpus (66 words, 0.16%) are tagged twice as many as they are in the male sub-corpus (25 words, 0.06%). Specifically, both the females and males prefer to describe the shape with nouns (F: *line, angle, cups, corner, loop*; M: *shape, square*) other than adjectives (F: *straight, flat*; M: *straight, 3D*). In the top 10 list, the females use nouns including *line* (9 times, 0.02%), *angle* (6 times, 0.01%), *cups* (4 times, 0.01%), *corner* (3 times, 0.01%), *loop* (3 times, 0.01%), and *spot* (twice), followed by four adjectives, namely, *flat* (3 times, 0.01%), *3d* (3 times, 0.01%) , *radical* (twice), and *sharp* (twice). The males use six nouns: *shape* (4 times, 0.01%), *square* (3 times, 0.01%), *nuggets* (once), *shaping* (once), *line* (once), and *circle* (once)

as well as three adjectives: *3D* (twice, 0.01%), *sharp* (twice, 0.01%) and *flat* (once). Meanwhile, both of the sub-corpora tag the word *straight* as an adverb instead of an adjective which is regarded as a mistag.

5. Getting and possession (A9+)

This is the only domain under the category of A: General and abstract terms as shown in Table 4.2 with a mixture of inflective form of four verbs, namely, *have*, *get*, *take* and *keep*. Both the females and males use the word *have* most significantly with a frequency of 152 times (0.37%) in the female sub-corpus and 117 times (0.30%) in the male sub-corpus. Additionally, for both the females and males, the rankings of different form of 'have' are the same, with *have* being most frequent, followed by *had* (females: 73 times, 0.18%; males: 45 times, 0.11%), *has* (females: 49 times, 0.12%; males:31 times, 0.08%), and *having* (females: 20 times, 0.05%; males: 13 times, 0.03%).

Interpretations with Related Research

Referring to two related research conducted by Schmid (2003) and Kaur (2009), the present study replicates some of their findings with some new perspectives as well. Schmid (2003: Section 6.1) tagged seven semantic domains as female preference, namely, *Clothing, Basic colours, Home, Food and drink, Body and health, Personal reference and personal relationships* and *Temporal deixis*. However, only four of them, i.e. *Home, Food and drink, Body and health, Personal reference and personal*

relationships, are related to the present study. Using Wmatrix, Kaur (2009) codes similar classifications of taggers from the girls' sub-corpus which is also the prototypical analytical framework of this study in order to investigate the gender differences in children's writing in Malaysia and the United Kingdom. More details of the correlation between the present study and the two significant research studies are compared and discussed in the following part.

1. People and Relationships

Combining the domains of Pronouns (Z8) and People: Female (S2.1) of the present study, it is evident that Malaysian females use more words such as *I* (most frequently used; F: 1392 times, 3.35 %; M: 1133 times, 2.89%), *she* (121 times, 0.29%; M: 19 times, %), *he* (F: 166 times, 0.40%; M: 94 times, 0.24%), and *girl* (F: 17 times, 0.04%; M: 5 times, 0.01%), which reconfirms Schmid's (2003) findings with the use of *we* (F: 283 times, 0.68%; M: 236 times, 0.60%) as an exception where the female frequency over-numbered that of the males.

Schmid (2003) tags the domain of *Personal references and personal relationships* (Schmid, 2009: Section 6.1.6) as a women's preponderance. In his results of personal references, it is found that women use more female names as well as words such as *she, girl, he* and *I* but men use more words including *man, we, men* and *people*. The use of

person is balanced between the two genders but is imbalanced in present study where males (44 times, 0.11%) use it more than the females (25 times, 0.06%).

Regarding personal relationships, the present study indicates that the Malaysian females (11 times, 0.03%) write more about their sons than the males (7 times) as noted in the domain of *Kin (S4)* whilst Schmid (2003) presents an opposite discovery that women use the word *son* slightly less than men. The two studies share the results that both women and females use more words that refer to *husband*, *sister*, *mother* in the S4 domain and *baby* in the domain of *Time: New and young (T3--)*. Conversely, the men and males use more wife-related words (e.g. *wife, my wife*). Although men are more prolific users of the word *parents* in Schmid's (2003) research, the present study reveals a slightly lower frequency in the male sub-corpus (8 times, 0.02%) than that of the female sub-corpus (9 times, 0.02%). Moreover, in the present study, the most popular family member is *baby* (32 times, 0.08%) in the female sub-corpus and *wife* (12 times, 0.03%) in the male sub-corpus whereas the women and men in Schmid's (2003) work talked about their *mums* most frequently.

Still on *people and relationships*, Kaur (2009: 162) groups *Pronouns (Z8)* and *People: Female (S2.1)* under this category. According to the present study, females write more about their same-sex friends which is also uncovered in Kaur's (2009) findings that girls made more references to the females owing to their exhaustive stories

deal with their girlfriends. Again, *I* is the most frequently-used pronouns in the two studies regardless of gender. The overuse of first-person pronouns (i.e. *I, my, we*) in the female weblogs from present study is also unfolded in Malaysian girls' writing (ibid, 2009).

One new discovery emerges in the use of the word *he* and *she* where the Malaysian females use *he* (166 times, 0.40%) to a greater extent when compared to the males whereas the Malaysian girls use *she* (929 times) more than *he* (508 times) when compared to the boys. The other one is that in the present study, more distinct gender differences are found in the Z8 domain (*Pronouns*) rather than the S2.1 domain (*People: Female*). On the contrary, Kaur (2009) discovers more obvious gender differences appeared in the S2.1 domain.

Additionally, the semantic domain S4 is the fourth entry in Table 4.3 of the present study, however, it is not of significant differences between Malaysian girls and boys but more salient differences emerged between British girls and boys in Kaur's (2009) findings. Moreover, the gender differences in the domain of *Kin* are evident in the present study but it is not so significant in the Malaysian children's corpora. Nevertheless, obvious gender differences are located in the British children's corpora where the British girls write more about their relatives (Kaur, 2009).

2. Body and Health

As it stands, the present study is consistent with Schmid's (2003) findings. Both of the studies share the overuse of *breast, hair, doctor, hand* and *eyes* in the women's sub-corpus and an equal frequency of *body* in the two corpora. Contradiction appears in the use of *ill* and *hands* which are women's preference in Schmid's (2003) study but turn out to be Malaysian male's predilection in the present study. In the present study, integrating the three semantic domains (*B3: Medicines and medical treatment; B1: Anatomy and physiology;* and *B2-: Disease*), the researcher finds that the word *asthma* (24 times, 0.06%) is most significantly used by Malaysian females whilst *sleep* (12 times, 0.03%) is used by Malaysian males to a greater extent in this category. However, according to Schmid (2003), *hand* is the most frequently used word in the men's and women's corpora under this category.

Concordance 4.10 Concordance line (females) of *asthma* (all 24 lines)

N			
1	fects on me. I was diagnosed with	*asthma*	since I was 4 year-old. But, It c
2	But, it can be said that my	*asthma*	is under control, which I can live
3	activity interrupted caused by an	*asthma*	attack, it does n't make me have t
4	ed to use Ventolin inhaler if mild	*asthma*	attack occurs. Then, everything b
5	italized. And, since that day my	*asthma*	symptoms become more frequent
6	t, due to lack of awareness about	*asthma*	, I was not very concerned about
7	l treatment delayed. Until 2009,	*asthma*	has changed my life completely.
8	etely. Several times I had severe	*asthma*	attacks and pneumonia. Since the
9	e doctor said it was caused by my	*asthma*	is getting worse. So, since that
10	I was hospitalized due to a severe	*asthma*	attack and intubated 4 times in th
11	h and that causes the frequency of	*asthma*	onset. Therefore, to help me get
12	very two weeks), I was free from	*asthma*	symptoms. Unfortunately, after t
13	after the treatment was stopped,	*asthma*	symptoms return. In fact, more o
14	treatment for patients with severe	*asthma*	. So, on 04.06.2012 I was admitte

Concordance 4.10, continued

Concordance 4.10 Concordance line (females) of *asthma* (all 24 lines)

15	er, after a few days at home, my	*asthma*	attacks came back and had to be
16	ng on my chest. It is only a mild	*asthma*	attack, but it could interfere me
17	d. On 3,.06.2012, I had a sudden	*asthma*	attack at 3 am. Since I am still b
18	t about 5 am once again I had an	*asthma*	attack and this time it's getting
19	IV and nebulizer. After that, an	*asthma*	attack gone until at 8am, again a
20	ttack gone until at 8am, again an	*asthma*	attack occur and lasts until 9:30 a
21	for a fifth time caused by severe	*asthma*	attack and admitted to the ICU.
22	o longer be hospitalized although	*asthma*	symptoms are still there. And, I
23	ly once my new admission due to	*asthma*	attack. It's been almost two mont
24	ts almost 3 weeks I was free from	*asthma*	attacks. In fact, I can breathe e

Concordance 4.11 Concordance line (males) of *sleep* (all 24 lines)

N			
1	h Alam until the nurse asked me to	*sleep*	. Thank you so much my friend
2	ame after the exam. I was about to	*sleep*	but was happy to see visitors! I
3	had deprived me of many nights of	*sleep*	. 2 years worth of sweat. Blood,
4	ime to rest. So, each time when I	*sleep*	around 12.30am, I would be ver
5	ery tired. Need to change that …	*sleep*	11/30p, after the e-meetings?
6	of food and have no proper place to	*sleep*	at night. They might do a kamik
7	e progress, but slow-Less morning	*sleep*	: Almost, but then afternoon or
8	ost, but then afternoon or evening	*sleep*	is inevitable –More focus to stu
9	1/2 of a real zombie due to lack of	*sleep*	plus a tiring day chasing for 'br
10	faded from the focus point. Hurray	*SLEEP*	SLEEP and plenty of SLEEP…
11	om the focus point. Hurray SLEEP	*SLEEP*	and plenty of SLEEP…After lu
12	urray SLEEP SLEEP and plenty of	*SLEEP*	… After lunch, napping before

3. Food

Parallel with Schmid's (2003) and Kaur's (2009) category of *Food and drink* (Schmid, 2003: Section 6.1.4; Kaur, 2009: 168), the results from the present study share similarities with the domain of *Food* in the two related studies.

According to Schmid (2003), women use more words than males in the category of *Food and drink* except the balanced usage of *beer* and men's overuse of *pizza* and *pint* than their women counterparts. In the present study, only two terms in the list of words are balanced, namely, *lunch* (F: 15 times, 0.04%; M: 15 times, 0.04%) and *restaurant* (F: 14 times, 0.03%; M: 13 times, 0.03%). The most frequently used word *dinner* in the women's sub-corpus (Schmid, 2003) is also significantly used in the female sub-corpus of the present study (21 times, 0.05%) whilst the most frequently used word *food* in the men's sub-corpus (Schmid, 2003) is most significantly used in the male sub-corpus of the present study (36 times, 0.09%). Specifically, both the women and men in Schmid's (2003) study discuss more about *dinner* rather than *breakfast* or *lunch*. However, the Malaysian male blogs make more references to *lunch* (15 times, 0.04%) whereas the females still concentrate on *dinner* (21 times, 0.05%) mostly.

As for the choice of particular food, rice, an Asian-featured staple food, predominates the Malaysian corpus with a frequency of 17 times in the female sub-corpus and 8 times in the male sub-corpus. However, Schmid's (2003) findings have no clues for any staple food. Instead, the women prefer *bread* (79.86%), *chocolate* (78.94%), and *cheese* (68.81%) whereas the men favour *bread* (47.58%), *cheese* (30.09%) and *eggs* (24.75%). This is probably because the Brown corpora and LOB corpora (Leech and Fallon, 1992) chosen by Schmid are discourse of American and British people who enjoy their own dietary habits.

Kaur (2009) also notes that Malaysian girls use more words under this category than boys. The girls write about what they like to eat as well as the family members' preparations for the food by using verbs such as *cooking* (20 times), *cook* (18 times), *bake* (7 times) and *baked* (twice). According to Kaur (2009), this indicates that the girls regard cooking experiences as an aspect of home life but the boys do not do so much. The present study is consistent with the findings in relation to the use of cooking-related verbs being more prevalent in the female sub-corpus. Therefore, the word *recipe* (21 times, 0.05%) and more words about ingredients such as *garlic* (9 times, 0.02%), *sauce* (8 times, 0.02%), *salt* (6 times, 0.01%), and *sugar* (4 times, 0.01%), are traced in the female sub-corpus.

Concordance 4.12 Concordance line (females) of *recipe* (all 21 lines)

N			
1	tion Information for 1/4 of Entire	*Recipe*	: Calories 350, Total fat 4.5g,
2	eate the Northstar Veggie Burger	*recipe*	out of thin air, and came up with
3	actural Northstar Veggie Burger	*recipe*	was published in a local newspaper
4	ed to the authentic sweet and sour	*recipe*	. INGRENDIENTS – Fish (whole o
5	ore weeks to go! xx, SS Chicken	*recipe*	My version Salam … I wanna shar
6	ersion Salam … I wanna share a	*recipe*	by accident which turned out surpri
7	e because I had also once tried a	*recipe*	from frend's blog similar to this
8	different, later I'll share that	*recipe*	too. So, I'll give the recipe to who
9	share that recipe too. So, this	*recipe*	will be given to someone who likes
10	Here's the steamed chicken's	*recipe*	. Ingredients: Chicken breast 150g
11	to share with you a rainbow cake	*recipe*	. To be honest I got this recipe fr
12	e recipe. To be honest I got this	*recipe*	from a good friend. Her name is Da
13	phne. Thanks a lot to her for the	*recipe*	and tips. She is really nice and g
14	and generous enough to share the	*recipe*	with me and she has given me gree
15	given me green light to share this	*recipe*	. Kan Daphne? .. hehe I was actual
16	ost more than 2 weeks. I got this	*recipe*	but I couldn't proceed myself to b
17	lots of friends requested for the	*recipe*	, I'll share it here. So, take a

Concordance 4.12, continued

Concordance 4.12 Concordance line (females) of *recipe* (all 21 lines)

18	he way, I modified a lil' bit the	*recipe*	though so that it'll be easy for y
19	r you to follow. Okay here's the	*recipe*	. Makeupless Cleopatra See! How t
20	and I decided to modify the basic	*recipe*	I found online with my own substit
21	Crunchy Supper Here is the basic	*recipe*	. I omitted a few items for my own

4. Parts of buildings (H2 tag)

Schmid (2003) sets the category of *Home* as a female's favoured topic (Schmid, 2003: Section 6.1.3). However, as the present study tags few items under this category, no significant gender differences are located in terms of related domains in her research. The only difference is that the present study notes that it is the males (6 times, 0.02%) not the females (4 times, 0.01%) who use the word '*door'* to a greater extent in this category. On the contrary, the women in Schmid's (2003) study use the word *door* more than their counterparts and it is also the most significantly used word within this domain.

Concordance 4.13 Concordance line (males) of *door* (all 6 lines)

N			
1	it still. Then he walk through the	*door*	and there he was … the person tha
2	next day, there's a knock on the	*door*	and standing before him a voluptuo
3	he next day there's a knock at the	*door*	and standing before him is the most
4	. n have me' Well, he's out the	*door*	after her like a shot. This girl is
5	he next day there's a knock at the	*door*	and when he opens it he finds a hug
6	english: lit. 5 fortunes approach	*door*	5 blessings bestowed upon your ho

5. Living creatures: animals, birds, etc. (L2)

In the present study, the theme within this category is not so pet-related. Instead, both the females and males link the animals to their food including *crabs* (7 times, 0.02%), *crab* (5 times, 0.02%), and *fish* (4 times, 0.01%) in the female sub-corpus and *fish* (3 times, 0.01%), *mussels* (twice, 0.01%) in the male sub-corpus. No wild animals are listed as the ten most frequently used words in the female sub-corpus, but the males, like the boys, record the living habits of some wildlife in their blogs such as *cows* (twice, 0.01%) and *snake* (twice, 0.01%).

According to Kaur (2009), the domain of *Living creatures: animals, birds, etc. (L2)* is the boys' indicator rather than the girls which is inconsistent with the present study. In her study, Kaur (2009) finds that Malaysian boys write about the animals they keep as pets (e.g. *fish*, 45 times; *dog*, 38 times; *cat*, 21 times; *bird*, 21 times) which results in the high frequency of the words: *pet* and *pets* (total, 50 times). The boys' texts also refer to wild animals (e.g. *frog*, 11 times; *goose*, 9 times; *bear*, 9 times; *eagle*, 8 times; *sharks*, 8 times; *monkey*, 7 times) by providing information about places of origin, animals' behaviours, their habits, breeding and eating habits as well as those animals they have encountered in zoos (e.g. *lion*, 15 times).

Interestingly, the more ferocious animals, namely *tiger* (twice) (Text 4.2) in the female weblogs and *cobra* (5 times, 0.01%) in the male weblogs reference the name of a

massage parlour and a movie respectively with only two concordance lines depicting the tiger show in a zoo. Differing from children's reliance on pet animals, the adults broaden their viewpoints and represent different aspects about animals. Therefore, the word *animal* (F: twice; M: twice, 0.01%) and *animals* (F: 4 times, 0.01%; M: twice, 0.01%) are more significantly used in the adults' corpora than that of the children's corpora.

Concordance 4.14 Concordance line (males) of *cobra* (all 5 lines)

N			
1	e one before this (G.I Joe Rise of	*cobra*	) . As for G.I Joe retaliation, I'd
2	cord (Marlan Wayans) of Rise of	*cobra*	? He is not dead on Rise of Cobra ..
3	Cobra? He is not dead on Rise of	*cobra*	.. but he seems lost. That's very od
4	w. He suppose be dead on Rise of	*cobra*	, after the last fight with snake ey
5	e eye under the Ice cap on rise of	*cobra*	.. surprisingly he is alive on retali

4.3.2 Key Semantic Domains in the Male Weblogs Sub-corpus

Given the male sub-corpus (see Chapter 3 for a further description of the corpora), Wmatrix tagged 39,222 words (95.13%) from the original texts, or 41,231 words. In total, the tagger generated 383 semantic categories which are more than that of the females (369 categories). Again, the researcher examined all the concordance lines in each of the semantic domains of the male sub-corpus to identify mistagged items, and removed the mistags from the analysis as they were not relevant based on the context of the word (see Appendix 3 for a list of mistags in the male sub-corpus). The distributions of the twenty-one top level of fields are shown in Table 4.3.

Table 4.3 Frequency Lists for Fields of Semantic Domains in Male Weblogs

Rank	**Fields of Semantic domains**	**Frequency**	**%**
1	A: General and abstract terms	85	22.19
2	N: numbers and measurement	51	13.32
3	X: psychological actions, states and processes	44	11.49
4	S: social actions, states and processes	43	11.23
5	T: Time	25	6.52
6	I: money and commerce in industry	18	4.70
7	O: substances, materials, objects and equipment	18	4.70
8	E: emotion	16	4.18
9	G: government and public	10	2.61
10	Q: language and communication	10	2.61
11	Z: names and grammar	10	2.61
12	B: The body and the individual	9	2.35
13	K: entertainment, sports and games	8	2.09
14	M: movement, location, travel and transport	8	2.09
15	W: world and environment	7	1.83
16	H: architecture, housing and the home	6	1.57
17	L: life and living things	6	1.57
18	F: food and farming	5	1.31
19	Y: science and technology	2	0.52
20	C: arts and crafts	1	0.26
21	P: Education	1	0.26

In this section, the male sub-corpus are analysed to examine the key semantic domains that are significant in the male sub-corpus in comparison with the female sub-corpus. Concerning the log-likelihood value, the field, namely, *C: arts and crafts*, failed to be tagged which indicates that there is no marked gender difference in that semantic category in relation to the female sub-corpus.

Wmatrix reveals twenty semantic domains (LL value > 10.83) that emerge from the analysis in which the adjusted frequencies and log-likelihood value are shown

(Table 4.4). It is worth noting that although the field *A: General and abstract* which predominates the semantic domain list also contains various types of tags (A11.1+, A3+, A6.1, A6.2+) when the males are compared to the females as identified by the log-likelihood value. The researcher conducts her analysis by taking the male sub-corpus as a basis. As an auxiliary, the female semantic domains are also examined where necessary.

Table 4.4 Key Semantic Domains in Male Weblogs Compared to Female Weblogs

Rank	**Tag**	**Semantic domain**	**Male weblogs**		**Female weblogs**		**LL value**
			Freq	***%***	***Freq***	***%***	
1	Z2	Geographical names	481	1.23	301	0.72	53.01
2	A11.1+	Important	82	0.21	26	0.06	33.87
3	W3	Geographical terms	91	0.23	32	0.08	33.06
4	N3.5	Measurement: Weight	46	0.12	9	0.02	29.44
5	G3	Warfare, defence and the army; weapons	74	0.19	28	0.07	24.29
6	M7	Places	177	0.45	105	0.25	23.04
7	Z5	Grammatical bin	10648	27.15	10592	25.47	21.60
8	I1.1+	Money: Affluence	21	0.05	2	0.00	19.43
9	K2	Music and related activities	68	0.17	29	0.07	18.50
10	I4	Industry	25	0.06	4	0.01	18.19
11	N5+	Quantities: many/much	205	0.52	143	0.34	15.03
12	I1.1	Money and pay	73	0.19	37	0.09	14.20
13	S1.1.3+	Participating	41	0.10	15	0.04	14.12
14	Y2	Information technology and computing	132	0.34	85	0.20	13.20
15	A3+	Existing	1384	3.53	1275	3.07	13.12
16	S7.3	Competition	9	0.02	0	0.00	13.01

Table 4.4, continued

Table 4.4 Key Semantic Domains in Male Weblogs Compared to Female Weblogs

17	A6.1	Comparing: Similar/different	18	0.05	3	0.01	12.78
18	S9	Religion and the supernatural	154	0.39	107	0.26	11.48
19	X2.6+	Expected	84	0.21	49	0.12	11.48
20	A6.2+	Comparing: Usual	94	0.24	57	0.14	11.45

As shown in Table 4.4, the strongest key semantic domain is *Geographical names (Z2)*. The most frequent domain in the male sub-corpus is *Grammatical bin (Z5)*, although this category is also frequently used by the females. As with the previous analysis, the researcher groups these domains based on emerging themes that relate to each other, bringing a total of five major groups from the table: *Geography and Location*; *Leisure*; *Money and Industry*; *Comparing and Evaluation*; *Measurement and Quantifiers*; *Social Actions and Process* and *Others*. Again, as in the female sub-corpus, the number of semantic domains under a group varies depending on the themes that link the domains together.

Geography and Location

Three semantic domains are grouped under this category: *Geographical names (Z2); Geographical terms (W3);* and *Places (M7).* According to previous studies (Mulac, Studley, and Blau, 1990; Mulac and Lundell, 1994), males make more references to *Quantity and Place* than females, while Johnstone (1993) also confirms the higher

frequency of *Place* as is consistent to the present study. Both the females and males share a high frequency in the use of *Chinese* and *Malaysia*. The males use the word *Chinese* most significantly 29 times (0.07%) and the word *Malaysia* 28 times (0.07%) whereas the females use *Malaysia* seventeen times and *Chinese* fifteen times respectively. The adjective *Chinese* frequently refers to the Chinese New Year festival, and relates to the fact that a group of the participants in the corpus are Malaysian Chinese and hence, they tend to record about this common celebration.

Concordance 4.15 Concordance line (males) of *Chinese* (13 of 29 lines)

N			
1	McDonald's Taiping During the	*Chinese*	New Year, I decided to bring my i
2	y and we all know that once the	*Chinese*	New Year celebrations draws eve
3	gogi Brothers, Mid Valle Happy	*Chinese*	New Year to all my fellow family
4	Instagram. Firecracker HAPPY	*CHINESE*	NEW YEAR 2013! I guess most
5	ith him every week. This year's	*Chinese*	New Year at the end of January 2
6	headed back to Ipoh to celebrate	*Chinese*	New Year. But I was greeted by a
7	just the 5kg program. Celebrate	*Chinese*	New Year with GABs C Fu, Lu a
8	whenever this happens, it means	*Chinese*	New Year is around the corner. A
9	n mind, GAB has put together a	*Chinese*	New Year offering that will rewar
10	Kelly, Ashley and Jane For this	*Chinese*	New Year, be sure to think of GA
11	runs every weekends (Sundays).	*Chinese*	New Year 2013 Third day of CN
12	lar terms in a year. However the	*Chinese*	New Year which is based on a lun
13	nake. We will be ushering in the	*Chinese*	New Year pretty soon and here is

As for other segments, the male sub-corpus represents more variety than the female. They use nouns such as *Hanoi* (20 times, 0.05%), *Myanmar* (12 times, 0.03%), *New York* (9 times, 0.02%), *Singapore* (8 times, 0.02%) and *KL* (8 times, 0.02%). They also use adjectives such as *Korean* (15 times, 0.04%), *Cantonese* (12 times, 0.03%) and

English (11 times, 0.03%). Since the males seem to be more concerned with international orientation, the females are more concentrated on domestic areas by using nouns such as *KL* (11 times, 0.03%), *Penang* (7 times, 0.02%) and *Ipoh* (5 times, 0.01%). The reason is probably that the majority of the females in the present study are married mothers who have to devote themselves completely to the family. Therefore, they choose the destinations at shorter distance.

It is worth noting that the males make references to historical events probably because the males tend to write about 'heroic actions' (Gambell and Hunter, 2000). For example, the word *British* (4 times, 0.01%) is linked with the invasion the country was once engaged in.

Concordance 4.16 Concordance line (males) of *British* (all 4 lines)

N			
1	eveloped in the 1880s when the	*British*	set up the tin mining industry, al
2	ited, (PCCL), which was under	*British*	control, had a 77-year lease to mi
3	ts that made up the realm of the	*British*	Raj. Meanwhile in the ancient cit
4	d further north, the influence of	*British*	India was felt again, especially I

Referring to *Geographical terms (W3)*, there are two main types. One is related to water area, whilst the other one is related to mountainous area. In the former category, the males use the words such as *island* (16 times, 0.04%), *beach* (6 times, 0.02%), *sea* (4 times, 0.01%), *coast* (3 times, 0.01%) and *wave* (3 times, 0.01%) while the females use the words such as *lake* (7 times, 0.02%), *beach* (3 times, 0.01%), *sea* (3 times,

0.01%), *pool* (3 times, 0.01%), *tides* (once), and *coast* (once). Regarding the latter category, the males use the words *valley* (3 times, 0.01%) and *Mount* (3 times, 0.01%) whereas the females use the words *hill* (twice), *slope* (twice) and *cliffs* (once). According to the statistics, the males use words with respect to Geographical terms (W3) approximately three times as much as the females (M: 91 times vs. F: 32 times), but the female descriptions seem more detailed than the male.

With regard to *Places (M7)*, which is a male's preferred domain (Mulac et al., 1990; Mulac and Lundell, 1994), it is consistent with Johnstone's (1993) work that males narrate places more than the females in spoken discourse. In the present study, the two most frequent words in the male and female sub-corpora in this tag are shared. These words are *place* and *city* with frequencies of 28 times (0.07%) and 18 times (0.05%) in the male sub-corpus, and 18 (0.04%) and 11 times (0.03%) in the female sub-corpora, respectively. In the top ten wordlist of this tag, they also share the use of *area* (M: 11 times, 0.03%; F: 5 times, 0.01%), *town* (M: 10 times, 0.03%; F: 3 times, 0.01%), *international* (M: 7 times, 0.02%; F: 3 times, 0.01%), *places* (M: 6 times, 0.02%; F: 10 times, 0.02%), and *countries* (M: 5 times, 0.01%; F: 4 times, 0.01%). Therefore, more gender similarities than differences are found in this domain.

Leisure

As Newman et al. (2008) state, males discuss more about leisure than their female counterparts. One semantic domain that falls under this group is *Music and related activities (K2)*. The terms in the male sub-corpus (0.17%) are more than that in the female sub-corpus (0.07%). The males are extremely interested in songs by using words such as *song* (5 times, 0.01%), *sang* (4 times, 0.01%), *songs* (3 times, 0.01%), *singing* (3 times, 0.01%) and *album* (3 times, 0.01%). The three most distinctive words are *hip-hop* (15 times, 0.04%), *pop* (twice, 0.01%) and *guitar* (twice, 0.01%) compared to the female sub-corpus. On the contrary, there are also words which occur distinctively in the female sub-corpus, namely, *jazz* (twice), *violin* (once) and *chorine* (once). Nevertheless, the most frequently-used word in the female sub-corpus is *song* (6 times, 0.01%).

Money and Industry

This category is made up of three semantic domains, i.e. Money: Affluence (I1.1+) Industry (I4) and Money and pay (I1.1). Nowson (2006) and Newman et al. (2008) also indicate that it is the males rather than females who talk more about money. Although the types in the domain of Money: Affluence (I1.1+) are not too many with 21 types in the male sub-corpus and two types in the female sub-corpus, the males still manifest multiple intentions than the females. Regarding the use of *prosperous* (M: twice, 0.01%; F: once), both of the genders use it in the context to make a wish.

Concordance 4.17 Concordance lines (males) of *prosperous* (all 2 lines)

N			
1	nd ceremonies for a blissful and	*prosperous*	new year. It marks the
2	3 hings ling4 english: business	*prosperous*	may you have a prosperous

Concordance 4.18 Concordance lines (females) of *prosperous* (only one line)

N			
1	then, looking forward towards a	*prosperous*	2013! Peace out. Grace It's a so

The lines in Concordance 4.17 and 4.18 indicate the similarities between the two genders with respect to a new year's resolution. The females also link the other type, *rich* (once), to make a wish. Meanwhile, the males also mention money making. Words such as *prosperity* (5 times, 0.01%), *wealth* (5 times, 0.01%), *bonus* (4 times, 0.01%), *wealthy* (once), *affluent* (once), *refund* (once) and *refunded* (once) are tagged from the text.

Significant differences arise in the domain: *Industry (I4)* as there are only four types in the female sub-corpus that are tagged: *workshops* (once), *mine* (once), *factory* (once), and *industry* (once). Upon examining the concordance lines, the word *mine* is used as a possessive pronoun other than a branch of industry. For the rest of the words in the female sub-corpus, they are entirely used in a blog to summarize the blogger's own experiences in 2012. Coincidentally, the lexical ambiguity of *mine* (5 times) is likewise noted in the male sub-corpus.

Concordance 4.19 Concordance line (males) of *mine* (8 of 13 lines)

N			
1	directed her to read an old post of	*mine*	about Dungun town in Trengganu
2	the worst part? THAT CD WAS	*mine*	, BLOODY HELL. I started to do a
3	were forgotten who we really are.	*Mine*	happened to be just right then, onl
4	that tasteless 14 years old self of	*mine*	. Prepared. I pressed play and forw
5	side which are very different from	*mine*	. Here they have drama queens,
6	t reads, ‘If I catch you, you’re	*mine*	. ’ He lost 31kgs that week. thanks
7	oney whenever your hand is not in	*mine*	Happy birthday darling I’ve no
8	linely moments by spendin more	*mine*	with you And I’d like to take away

In general, *mine* in Concordance 4.19 are completely served as a possessive pronoun. The other five are linked with a historical story with one being verbal form and four being substantial forms. The males demonstrate their single-minded focus on mining industry by using related words such as *mining industry* (twice, 0.01%), *mining* (twice, 01%) and *pit mines* (once) apart from *mine* (5 times, 0.01%). They also mention about other industries incorporate *aviation industry* (once) and *IT industries* (once).

Concerning the domain of *Money and Pay (I1.1)*, the males are interested in current assets by using words including *profit* (8 times, 0.02%), *investment* (5 times, 0.01%) and *investments* (twice, 0.01%) whilst the females care more about the fixed assets with the words such as *save* (10 times, 0.02%), *saving* (once), *saved* (once) and *savings* (once) instead. Although the females use the word *investment* four times, three of them refer to a chat rather than a business activity they engage in.

Text 4.3 Script Female 83

We also discussed about the Geneva Gold ***investment***.

"You think the money you lend to Botak is more risky or the Geneva Gold ***investment***?" I asked.

"Of course Geneva Gold ***investment*** is more risky as we do not know what they do with our money but I know what is Botak doing and his business model, he lend the money to the Indonesian labourers and keep their passports. He charged them 40% interest per month and give me 4% because he said he has to do all the donkey work like go after the debtors if they fail to pay their debts. It is some sort like micro-financing."

Moreover, the males provide more terminologies such as *hedge fund* (twice, 0.01%), *mutual fund* (twice, 0.01%), *dividends* (twice, 0.01%) and *fund* (once) which are hardly noted in the female sub-corpus, whereas the females prefer more general and common words including *pay* (twice), *tax* (once), *income* (once) and *afford* (once) which are all embodied in the male sub-corpus.

Comparing and Evaluation

There are three semantic domains under this category, namely, *Comparing: Similar/different (A6.1)*; *Comparing: Usual (A6.2+)*; and *Important (A11.1+)*. The number of words pertaining to the domain *Comparing: Similar/different (A6.1)* is slightly small in both sub-corpora. One word shared by both gender is *compared* with a frequency of 8 times (0.02%) in the male sub-corpus and twice as many in the female sub-corpus. All the ten concordance lines are life-related with multiple topics.

Concordance 4.20 Concordance line (males) of *compared* (all 8 lines)

N			
1	of the snake plant is very strong	*compared*	to the flimsy money plant leaves.
2	r earnings per share of 41.19 sen	*compared*	with 29.98 sen previously. The gr
3	positive side, it is less oil as	*compared*	to its Penang counterpart and coul
4	I'm no longer a regular blogger	*compared*	to previous years before. Well,
5	s. She's rather nice and stable	*compared*	what I faced in the crappy past wh
6	. It was a less well island	*compared*	to other islands in Terengganu
7	public holidays in Singapore (as	*compared*	to Malaysia, our country has most
8	its modern equipment helps a lot	*compared*	to those old school clinics. And

As shown in Concordance 4.20, line 1 is about comparison between plants, line 2 is about money, line 3 refers to food while line 4 and 5 are related to blogging. Line 6 and 7 are about travel and holiday while line 8 talks about medical treatment. Apart from *compared*, the other type used by the females is *balance out* (once). One woman wrote in her blog: "I'm not a fan of sweet foods as a main dish, unless it's balanced with savoury and spicy flavours. In this dish, the honey, soy sauce, rice vinegar and sriracha complement each other and balance out the sweetness". Again, this description is related to food and how to keep healthy.

The last entry of the list is the domain of *Comparing: The words usual (A6.2+)*. *Usually* (M: 10 times, 0.03%; F: 3 times, 0.01%), *common* (M: 10 times, 0.03%; F: 7 times, 0.02%), *normal* (M: 4 times, 0.01%; F: 7 times, 0.02%), and *tend* (M: 5 times, 0.01%; F: 8 times, 0.02%) emerge in both sub-corpora and the males and females both use the words *wont* and *usual* 4 times respectively. In the female sub-corpus, the word *usually* (3 times, 0.01%) is mainly linked with the bloggers' preference in food and

garment while in the male sub-corpus, the word *usually*, also known as the most frequently used word in this domain, only one line refers to the blogger's travel plan compared to what he has always done before whilst other lines are not so self-centred.

Under the domain of *Important (A11.1+)*, three times more occurrences are tagged in the male sub-corpus than that in the female sub-corpus. Being the most significantly used word, *important* (17 times, 0.04%) is principally used in its comparative and superlative degrees in the male sub-corpus while solely used once in the female sub-corpus.

Concordance 4.21 Concordance line (males) of *important* (10 of 17 lines)

N			
1	than June sometime. But more	*important*	than "When" will be the question
2	of thinking about other more	*important*	things. In fact, Grampa suspecte
3	learned that there are other much	*important*	aspects in a marriage. Your favor
4	eally matter to em. What's more	*important*	. Prepared. I pressed play and
5	all the new things and most	*important*	. Here they have drama queens,
6	going to wear. Is not life more	*important*	. ' He lost 31kgs that week.
7	than food? Is not the body more	*important*	Happy birthday darling Ive no
8	feeds them! Are you not more	*important*	with you And Id like to take
9	and making our time the most	*important*	commodity. It will be nice to liv
10	knowledge and financial. Most	*important*	point to remember is to help

Concordance 4.22 Concordance line (females) of *important* (1 of 3 lines)

N			
1	to make it and family is the most	*important*	thing to me, so I just WANT to

The females, by contrast, have a passion for using phases such as *means a lot* (once), *well known* (once) and *make all the difference* (once). They even use the colloquial word *biggies* once while the males use more formal words: *value* (5 times, 0.01%), *value* (3 times, 0.01%), *significant* (3 times, 0.01%), *upgraded* (twice, 0.01%) and *upgrade* (twice, 0.01%).

Measurement and Quantifiers

The domains *Measurement: Weight (N3.5)* and *Quantities: many/much (N5+)* are grouped under this category. The fourth entry is *Measurement: Weight (N3.5)* which is used six times as frequently in the male sub-corpus (0.12%) relative to the female sub-corpus (0.02%). This is probably because males are keen on heroic behaviour while the demand for physical strength never ends (Quicke, 2012).The most frequent word in this domain is *weight* which occurs eleven times in the male sub-corpus and twice in the female sub-corpus. However, referring to the unit of measurement, a difference is noticed between these two genders. Specifically, the males prefer to weigh by *Kg* whereas the females are in favour of *Pound*.

Moreover, there are more figures in the male sub-corpus than in the female sub-corpus with the words being *130 kg* (5 times, 0.01%), *5 kg* (3 times, 0.01%), *45kg* (twice, 0.01%), *1 kg* (twice, 0.01%) and *2 kg* (once) in the former and merely *21b* (once), *2-31bs* (once) and *1bs* (once) in the latter. Furthermore, both genders seem to be

concerned about popular issue, namely, losing weight by using the word *overweight* twice in the male sub-corpus, the word *heaviness* once in the female sub-corpus and the word *pressure* (M: 3 times, 0.01%; F: once) in both sub-corpora.

According to Mulac et al. (1990), Mulac and Lundell, (1994), Quantity is a domain favoured by men. In the present study, it is manifested that the males and females share the rankings of words including *much* (M: 33 times, 0.08%; F: 29 times, 0.07%), *enough* (M: 32 times, 0.08%; F: 24 times, 0.06%), *a lot* (M: 14 times, 0.04%; F: 16 times, 0.04%), *lots* (M: 10 times, 0.03%; F: 9 times, 0.02%) and *increase* (M: 7 times, 0.02%; F: 3 times, 0.01%).

In particular, the word *increase* is used as a verb form rather than a noun form in both sub-corpora (only used as noun once respectively), but the two lines in the female sub-corpus merely talk about financial issues, whereas the males broaden into health (line 1 and 5), architecture (line 2), transportation (line 3), financial issue (line 4), and social activity (line 6) as shown in Concordance 4.23. This is consistent with the findings of Newman et al. (2008) who state that males prefer to write about their current concerns.

Concordance 4.23 Concordance line (males) of *increase* (6 of 7 lines)

N			
1	for giving me advise on how to	*increase*	the blood count. Thank you Kus!
2	now. One of the way we could	*increase*	the spaces are to have ‘ built-ins
3	ehow or rather, the size of cars	*increase*	… over the years. If we are sin
4	achieve our goals (financial)	*increase*	our knowledge (financial)… then
5	increases your BMR as you	*increase*	your lean muscle mass which
6	ONG can utilize this technique to	*increase*	his/her popularity graph. It is a

Social Actions and Process

Three semantic domains are under analysis in this category, namely, *Participating (S1.1.3+), Competition (S7.3)* and *Religion and the supernatural (S9)*. In the domain of *Participating (S1.1.3+),* the males (41 terms, 0.10%) use nearly three times more terms than the females (15 times, 0.04%). Taking the two most frequently used words as examples, the males use the word *conference* (12 times, 0.03%) much more than the females (twice) with different foci. In the male sub-corpus, aside from three concordance lines dealing with routine of conference, all the other 9 concordance lines refer to the writing of a conference paper (Concordance 4.24). The two concordance lines in the female sub-corpus, however, refer to the conference speaker and the schedule of a conference.

Concordance 4.24 Concordance line (males) of *conference* (9 of 12 lines)

N			
1	active and just finished my first	*conference*	paper after 5.5 months being a
2	took almost 2 weeks to write my	*conference*	paper together with the
3	difficult it is writing a technical	*conference*	paper. I remember when I did my
4	long did they take to publish a	*conference*	paper and most of them said after
5	should be planning to write two	*conference*	papers in the first 12 months of
6	Anyway I could do a second	*conference*	paper if I want by taking a “saf
7	hs then this could be the second	*conference*	paper of my first year PhD. But
8	But I was busy writing my	*conference*	paper, I did put lots of effort
9	rtain because it depends with my	*conference*	paper. One wonderful news is

Concordance 4.25 Concordance line (males) of *meeting* (8 of 9 lines)

N			
1	my work in the academic	*meeting*	for TWICE. One is in Hokkaido
2	work presented in the academic	*meeting*	on February.” Which shows that
3	my work in the future academic	*meeting*	. A honor for every master
4	my presentation in the academic	*meeting*	. I stay numb for maybe a few
5	the way, it is the same academic	*meeting*	I presented my work last year
6	Submission for Academic	*Meeting*	2013 January 28th Thesis
7	2013 February 1st Academic	*Meeting*	2013 February 14-15th Final
8	my work to the academic	*meeting*	in my first year of master which

The other one word, *meeting*, is used in the male sub-corpus (9 times, 0.02%) to a greater extent compared to that of the female sub-corpus (twice). The two concordance lines in the female sub-corpus refer to meeting people when traveling while 8 of 9 concordance lines in the male sub-corpus record link to academic meeting as shown in Concordance 4.25.

Apart from these academic participation, the males also refer to some leisure gathering by using words such as *reunion* (6 times, 0.02%) and *parties* (twice, 0.01%)

while the females use more verb phases including *met up* (once) and *joined in* (once) to depict their entertainment. More verb-oriented style of female writing is also noted in previous studies (Newman et al., 2008; Schultz, 2013).

As there is no term tagged in the female sub-corpus, all the terms in the *S7.3* tag *Competition* originate from the male sub-corpus. The males use six terms in total to represent different genres of competition. The words *contest* (3 times, 0.01%) and *opponent* (once) refer to competition in conversations with the word *finalists* (twice, 0.01%) deals with TV competitions. The words *tournament* (once) and *tournaments* (once) are related to games while *rivalries* (once) make references to education.

In terms of the domain *Religion and the supernatural (S9)*, the females and males share the usage of some words including *God* (M: 19 times, 0.05%; F: 7 times, 0.02%), *soul* (M: 11 times, 0.03%; F: 4 times, 0.01%), *Christmas* (M: 6 times, 0.02%; F: 10 times, 0.02), and *temples* (M: 4 times, 0.01%; F: twice). Regarding the use of the typical religious word *God*, 16 of 19 concordance lines are related to its religious meaning, i.e. the creator or ruler of the world they worshipped in individual religious faith. The other three concordance lines refer to ‘pray for something good’ (twice) and ‘exclamation of

emotion' (once). On the other hand, one female uses one phrase "for God sake" to show her *anger* towards the *Bawani event*[13] (Text 4.4).

Text 4.4 Script Female 3

We should not just be contented and not want to work further, NO, we should strive harder for a better future for our later generation and ***for God sake***, stop talking about animals, what does the animals have to do with our election system and our educational system, the ministry should send a better representative to chair the meeting, the rebuttal is just ridiculous!!

Although both the males and females use the word *Christmas*, their references are different. As for the males, this word in all the six concordance lines simply serves as a time indicator of that particular day (December twenty-fifth), whereas the females refer to celebration (line 2), shopping (line 6) and presents (line 1, 3, 4, 5) as shown in Concordance 4.26.

Concordance 4.26 Concordance line (females) of *Christmas* (6 of 10 lines)

N			
1	y miss A hope you all has a great	*Christmas*	so far. Mine was okay too, and as
2	first, I thought it's a typical	*Christmas*	ornament, like the ones you will
3	to abby. Jasamu dikenang. hehe.	*Christmas*	edition-Doraemon ochado Card
4	card design is set for 3 editions:	*Christmas*	, Valentine's and Ochado. And fo
5	tines' and Ochado. And for this	*Christmas*	edition card, it is Doraemon. Wo
6	just to window shop at the	*Christmas*	Sale that was going on. And little

[13] Bawani is an undergraduate who studies law in Universiti Utara Malaysia and issued about the free education in Malaysia against the forum speaker, Sharifah Zohra. But she was forcibly interrupted by the presenter with the famous words "Listen! Listen! Listen!" and her microphone was also taken away by Sharifah for avoid further questionings in public.

Differing from the typical religious indication of the word *God*, the word *temple* with a frequency of four times in the male sub-corpus and twice in the female sub-corpus, is related to travel without exceptions. On one side, females refer to two travel destinations, namely, Pink City Jaipur in India (once) and the resorts of Thailand (once). On the other hand, the males point to the historic sites in Hanoi (twice), Cambodia (once) and Indonesia (once).

Other distinct words are places-related, e.g. *church* (8 times, 0.02%), spirit-related, e.g. *Holy Spirit* (5 times, 0.01%), *spirit* (5 times, 0.01%), *holy* (3 times, 0.01%), and belief-related, e.g. *Tao* (5 times, 0.01%) and *Islam* (3 times, 0.01%) in the male sub-corpus and *witch* (5 times, 0.01%), *pray* (4 times, 0.01%), *hell* (3 times, 0.01%), *heaven* (3 times, 0.01%), *sacrificing* (3 times, 0.01%) and *Muslim* (twice) in the female sub-corpus. References to Islam and Muslim owe to the fact that Malaysia has a majority of Muslim population as revealed in the International Religious Freedom Report for 2011.

> According to 2010 census figures, 61.3 percent of the population practices Islam; 19.8 percent Buddhism; 9.2 percent Christianity; 6.3 percent Hinduism; and 1.3 percent Confucianism, Taoism, and other traditional Chinese religions… Ethnic Malay Muslims account for approximately 55 percent of the population.

Others

The researcher has grouped the rest of the semantic domains under this category as they do not relate to the other categories. They are the *G3* tag *Warfare, defence and the army*; *weapons*; the *Z5* tag *Grammatical bin*; the *Y2* tag *Information technology and computing*; the *A3+* tag *Existing*; and the *X2.6+* tag *Expected*.

1. Warfare, defence and the army; weapons (G3)

The fifth most significant domain used by males is *Warfare, defence, and the army; weapons (G3)*. Statistics reveal that the males use the relevant words twice as many as the females (M: 74 types; F: 28 types). This is also revealed in Peterson's (2002) work that men prefer to write more about violence and Flynn (1983) identifies gun control and nuclear power as typical topics among the men.

Both the males and females mention about shooting-related words such as *shot* (M: 3 times, 0.01%; F: 5 times, 0.01%), *shoot* (M: twice, 0.01%) and *shooting* (M: twice, 0.01%; F: 4 times, 0.01%). As the females only record these words *bullet* (3 times, 0.01%) and *gun(s)* (3 times, 0.01%), the males move their attention to the whole army thing by using words like *army* (9 times, 0.02%), *marines* (7 times, 0.02%), *guardsmen* (twice, 0.01%), *armor* (twice, 0.01%), *barrack* (twice, 0.01%) and *squad* (once) which never appear in the female sub-corpus.

In comparison to the females, the males also use more words that refer to aggression, especially weapons and artillery used in conflicts and war: *guns* (5 times, 0.01%), *WWII* (3 times, 0.01%), *sword* (twice, 0.01%), *bombs* (twice, 0.01%), *weapons* (once), *raided* (once), *firepower* (once), *machine guns* (once), *grenade* (once), and *war* (once). The female sub-corpus is restricted to some extent to convey the people involved. For example, one woman uses the term *spy* once to talk about an impression of people in a place.

Concordance 4.27 Concordance line (females) of *spy* (only one line)

N			
1	Vegas is an excellent backdrop for a	*spy*	film and a playground reserved for th

2. Grammatical bin (Z5)

The Z5 tag has the seventh most significant difference (LL value = 21.92) indicating an overuse of the semantic domain in the male sub-corpus in which the males use a relatively higher number of words (27.15%) than the females (25.47%). This includes pronouns such as *the* (1859 times) and *a* (777 times); preposition: *to* (1103 times), *of* (758 times), *in* (518 times), *for* (423 times), and *with* (275 times) and conjunctions: *and* (1123 times) and *but* (232 times). Surprisingly, the females share the ten most significantly used words with the males with *the* (1623 times) being the most frequently used one.

3. Information technology and computing (Y2)

Flekova and Gurevych (2013) indicate that males talk more about computer than females. Even in the teenager groups, 60% of the boys are reported to discuss more about science-related inventions (Levine and Geldman-Caspar, 1996). Owing to the fact that the data of this study is from weblogs, blog-related terms dominate the frequency list of this domain. Both the males and females use the word *blog* most frequently with a frequency of 38 times (0.10%) in the male sub-corpus and 40 times (0.10%) in the female sub-corpus. The majority of the word reflects the blogger's blogging experiences himself/herself as 21 of 32 concordance lines in the male sub-corpus and 28 of 40 concordance lines in the female sub-corpus are phased with first personal pronouns. This is in accordance with previous studies (Argamon et al., 2003; Nowson, 2006; Newman et al., 2008; Yu, 2013) that females tend to use more first personal pronouns and write about themselves. Other words shared by the two gender groups are *internet* (M: 5 times, 0.01%; F: 4 times, 0.01%), *blogging* (M: 4 times, 0.01%; F: 3 times, 0.01%) and *bloggers* (M: 4 times, 0.01%; F: 5 times, 0.01%).

Furthermore, the males use the words 'website' and 'computer' by using words including *website* (9 times, 0.02%), *websites* (4 times, 0.01), *computer* (3 times, 0.01%) and *computers* (once) whilst the females mention about the *computer(s)* (twice in total) as well as *screen(s)* (totally 5 times) and *laptop* (twice).

4. Existing (A3+ tag)

The semantic domain *A3+* encompasses all the forms of the verb 'to be' and is more frequent in the male sub-corpus compared to the female's sub-corpus. This is mainly because of the high frequency of the verb *is* (M: 487 times, 1.24%; F: 365 times, 0.88%). However, when relative frequencies are taken into account, it is clear that both males and females use the forms of 'to be' a lot (M: 1384 terms, 3.53%; F: 1275 terms, 3.07%).

Additionally, for both males and females, the ranking of the different forms of 'to be' are the same, with *is* being most frequent, followed by *was* (M: 216 times, 0.55%; F: 254 times, 0.61%), *be* (M: 181 times, 0.46%; F: 148 times, 0.36%), *are* (M: 171 times, 0.44%; F: 127 times, 0.31%) and *'s* (M: 115 times, 0.29%; F: 116 times, 0.28%). It is difficult to summarise how these words are used as they are so frequent, but it is perhaps notable that males are more likely than females to reference states of being, as supposed to say, verbs like 'do' or 'have'. It is also the case that words in this category tend to operate alongside other categories.

5. Expected (X2.6+ tag)

In total, five items are tagged both in the male and female sub-corpora respectively. The words *hope* (M: 33 times, 0.08%; F: 15 times, 0.04%), and *expect* (M: 9 times, 0.02%; F: 8 times, 0.02%) are also the two most frequently occurring words in the

female and male sub-corpora. The other three terms are *hopefully* (M: 6 times, 0.02%; F: 3 times, 0.01%), *looking forward* (M: 5 times, 0.01%; F: 4 times, 0.01%), and *expected* (M: 5 times, 0.01%; F: 6 times, 0.01%). Regarding the concordance lines of the word *hope*, both the males and females use the phase *I hope* (M: 27 times, F: 9 times) significantly with only one line in each sub-corpus that the word is used in its noun form. As the males use the variants of 'hope', 'expect' and 'look forward', the females, apart from the three basic terms, also use the verbs *anticipated* (once) and *foresee* (once) to describe their own psychological activities.

Concordance 4.28 Concordance line (females) of *anticipated* and *foresee* (all 2 lines)

N			
1	of parenting- one that I had not	*anticipated*	5. Pushes the envelope .. well
2	ey are no longer available. I can	*foresee*	myself drowning in caffeine addic

Interpretations with Related Research

Taking men's writings as a prototype, Schmid (2003: Section 6.2) establishes seven domains as the males' stylistic features, namely, *Swearwords*; *Car and traffic*; *Work*; *Computing*; *Sports*; and *Public affairs*. Again, the present study merely replicates one domain, i.e. *Computing*. However, in Kaur's (2009) findings, more similar categories emerge in the present study as well.

1. Geography and Location

The following concordance lines present the word *Japanese* in context.

Concordance 4.29 Concordance line (males) of *Japanese* (all 3 lines)

N			
1	experiencing now? Or is this the	*Japanese*	lab culture that has affected me n
2	ke myself familiar with how the	*Japanese*	do things and think how to work w
3	m. So, I need to think like the	*Japanese*	and I need to improve myself to be

Concordance 4.30 Concordance line (females) of *Japanese* (only one line)

N			
1	y for you: All the gifts. Those	*Japanese*	junk food from my mom. My mo

In the present study, the word *Japanese* (3 times, 0.01%) refers to *Japanese culture* (once), *Japanese people's thoughts* (once) and *behaviour* (once) in the male sub-corpus along with *Japanese junk food* (once) in the female sub-corpus. Referring to Kaur's (2009) work, she also found that the Malaysian boys making references to historical events. However, the words *Japanese* (11 times), *American* (10 times), and *Portuguese* (4 times) relate to wars that these countries have engaged in. In opposition to the boys, the Malaysian girls use the word *Japanese* for a *Japanese Garden* (twice), a *Japanese tea house* (once), *Japanese food* (twice), and a *Japanese cartoon* (once).

Another similarity between the two studies lays in the overuse of the words 'Chinese' and 'Malaysia' with frequencies of 21, 16 and 11 times in the boys' sub-corpus and 43, 19 and 13 times in the girls' sub-corpus respectively. Both the adults' corpora (present study) and children' corpus (Kaur's study) use the adjective *Chinese*

frequently to refer to the 'Chinese New Year' festival which is a common celebration for the Malaysian Chinese and Chinese people.

One more point in common between the two studies grounds in intrinsic denotation of a geographical name which refers to travel sites such as *Sydney* (8 times) and *India* (7 times) in the boys' sub-corpus, and *Penang* (9 times) and *Singapore* (8 times) in the girls' sub-corpus with no significant gender difference in the choices of international destinations[14] and domestic destinations[15].

One distinction between the two studies emerges from the indicator of 'football'. There is no concordance which indicates this specific feature in neither the male nor the female sub-corpus in the present study. However, in Kaur (2009)'s work, she identified that the Malaysian boys make references to football teams by words such as *Arsenal* (11 times), *Brazil* (7 times), *Madrid* (6 times), *Liverpool* (4 times), and *Chelsea* (once); even the Malaysian girls make references to one football club, namely, *Liverpool* (twice).

[14] As indicated in the present study that the Malaysian males prefer to travel abroad (e.g. Hanoi, Korean, Myanmar and New York).
[15] As revealed in the present study that the Malaysian females favour domestic destinations (e.g. KL, Penang and Istanbul).

2. Leisure

Both of the corpora (Malaysian adults and children) share a large proportion of words associated with 'singing' including *songs, singing, sing, singer* and *song*. The males in the present study and boys in Kaur's (2009) study write about the same musical instrument, namely, *guitar*, twice respectively. According to Kaur (2009), the *K2* domain was the fourth most significant domain in the Malaysian girls' texts compared to the boys. While the females in the present study only mention about *violin* (once) and *drum* (once), the girls use the word *piano* (22 times) most frequently (Kaur 2009). Other hobbies of men are indicated in the corpus, too. Though the males refer to *hip-hop* (15 times, 0.04%) and boys refer to *rapper* (twice), these two words are never represented in the female or girls' sub-corpus.

3. Warfare, defence and the army; weapons (G3 tag)

Concordance 4.31 Concordance line (males) of *army* (all 9 lines)

N			
1	a regular base like the rest of the	*army*	has. I'll use the 40mm base for Ca
2	a budget trip. Do You Think The	*Army*	Should Handle Lahad Datu? From
3	us post many people feel that the	*Army*	should step in to handle this Sulu a
4	ffair in Lahad Datu. I think by "	*Army*	" it is also meant the Army, Air Fo
5	k by "Army" it is also meant the	*Army*	, Air Force and Navy. This is not a
6	rvey asks if we should call in the	*Army*	. Please vote quickly. If there are
7	h the Al Maunah heist it was the	*Army*	that handle the situation. Mainly
8	p and stolen weaponry from the	*Army*	. The Maunah was only a criminal
9	was only a criminal gang yet the	*Army*	stepped in. I think that was also b

In line with previous discussions, the adult males in the present study and boys in Kaur's (2009) research write about this topic. Kaur (2009) classifies this domain under the category of Aggression and adventure (Kaur, 2009: 173). Compared to the boys, the males in the present study depict armed forces with two most significant words, namely, *army* (9 times, 0.02%) and *marines* (7 times, 0.02%). However, the boys use more words to refer to weapons and artillery such as *sword* (20 times) (most frequent word), *weapons* (3 times), *machine gun* (3 times), *laser gun* (3 times), *guns* (3 times), *grenades* (3 times), and *bullets* (twice).

Additionally, three shooting-related words (*shot, shoot* and *shooting*) are notified in the present study which are inconsistent with Kaur's (2009) finding where the words *shot* (7 times) (second most frequently occurring words in this domain) and *shoot* (twice) are identified under this domain. The girls convey the same information as the females that instead of writing about those aggression, their foci are in the people that are involved in such acts such as *spies* (9 times) (also found in present study), *warrior* (6 times), *warrior* (5 times), *soldiers* (4 times), and *officer* (3 times).

4. Information technology and computing (Y2 tag)

Concordance 4.32 Concordance line (males) of *computer* (all 3 lines)

N			
1	it only takes several clicks on a	*computer*	to find all the answers. With infor
2	generally, in no where the words	*computer*	pops up in the definition of IT fro
3	ip-hop arena without leaving my	*computer*	chair. Kinda explains all the extent

While further exploring the corpora, the word *computer* occurs 3 times (0.01%) in the male sub-corpus and once in the female sub-corpus. Its plural form, *computers*, is only used once in the Malaysian adults' corpus. It is not until December 1997 that Jorn Barger coined the term *Weblog*. Later, the British physicist Tim Berners-Lee, who is also the creator of the World Wide Web, built the very first weblog on the website http://info.cern.ch/ (Singh and Shahid, 2006). However, the BNC conversational corpora was established at the end of the 1980s and the beginning of the 1990s when the majority of the people did not have too much contact with computer let alone weblog (Schmid, 2003: Section 6.2.4). Consequently, there was merely low frequency of words in the list of Schmid's (2003) work which were all basic terms related to computer such as *printer, desktop* and *Windows*. But still, the men in Schmid's (2003) study used more words in this domain than their counterparts with the word *computer* being the most popular word in the corpora.

5. Existing (A3+ tag)

Both the two studies share similarity in word frequency and ranking to some extent, except that the present study informs that it is the male not the female who use more words such as *is*, *be* and *are*. Moreover, the two studies concur with the usage of *was* where the Malaysian females and girls use the word more than their counterparts. Contrary to the present study, Kaur (2009) finds out that Malaysian girls (3494 terms, 4.93%) use much more terms under this domain than boys (1792 terms, 4.22%).

4.4 Analysis of Key Parts of Speech (POS)

As an auxiliary analysis, the gender difference in POS is also examined by the researcher. Adopting the same methodology in Section 4.3, the parts of speech with LL value over 10.83 are listed and analysed with classification of female indicator and 'male indicator' (Biber et al., 1998). The next step is to analysis the key parts of speech (POS) in female sub-corpus (Section 4.4.1) and the key parts of speech in male sub-corpus (Section 4.4.2) in sequence. As the findings of relevant studies are extremely similar and also consider the auxiliary function of this part, the interpretation of related research with the present study are mingled in context.

4.4.1 Key Parts of Speech in the Female Weblogs Sub-corpus

Table 4.5 Frequency Lists for Key Parts of Speech in Female Weblogs

Rank	**Tag**	**Key Part of Speech**	**Frequency**	**%**
1	NN1	singular common noun	6117	14.71
2	JJ	general adjective	2496	6.00
3	II	general preposition	2038	4.90
4	RR	general adverb	1737	4.18
5	AT	article	1690	4.06
6	NN2	plural common noun	1540	3.70
7	VVI	infinitive	1473	3.54
8	VV0	base form of lexical verb	1358	3.27
9	PPIS1	1st person sing. subjective personal pronoun (*I*)	1315	3.16
10	CC	coordinating conjunction	1311	3.15

As shown in Table 4.5, females use *nouns* (e.g. *NN1*, *NN2*) most frequently followed by *adjectives (JJ)*, *prepositions (II)*, *adverbs (RR)*, *articles (AT)*, *pronouns*

(PPIS1) and *conjunctions (CC)*. The base form of *verbs (VV0)* and *infinitives (VVI)* are also noticeable in the female sub-corpus.

Table 4.6 Key POS in Female Weblogs Compared to Male Weblogs

Rank	**Tag**	**Part-of-speech**	**Female weblogs**		**Male weblogs**		**LL value**
			Freq	***%***	***Freq***	***%***	
1	PPHS1	third person sing. subjective personal pronoun (he, she)	287	0.69	113	0.29	68.44
2	PPHO1	third person sing. objective personal pronoun (him, her)	117	0.28	38	0.10	37.73
3	UH	interjection	195	0.47	89	0.23	34.58
4	VVD	past tense of lexical verb	930	2.24	673	1.72	27.72
5	VV0	base form of lexical verb	1358	3.27	1077	2.75	18.14
6	PPIS1	first person sing. subjective personal pronoun (I)	1315	3.16	1041	2.65	17.92
7	RR	general adverb	1732	4.18	1435	3.66	13.84
8	VBM	am	166	0.40	102	0.26	11.92

As shown in Table 4.6 which includes the POS with a LL value > 10.83, a large proportion of pronouns are encompassed in the female sub-corpus as it is a predominant female indicator (Biber et al. 1998; Koppel et al., 2002). Overall, the most significant gender difference occurs with the use of *he* (F: 166 times; M: 94 times) and *she* (F: 121 times, M: 19 times) which are also highlighted by Argamon et al. (2003). In general, females use third person (*he, she, him, her*) and first person (*I*) pronouns more

frequently than their counterparts which is consistent with previous studies (e.g. Biber et al., 1998; Newman et al., 2008; Yu, 2013). Also noted is that they use general adverbs (1737 times) slightly more than the males (1435 times) as Newman et al. (2008) has also testified.

Moreover, the POS tagger of *interjection (UH)* revealed that the females use interjections twice as many as the males which is in accordance with relevant studies (Bednarek, 2009; Ali and Aslam, 2012; Rangel and Rosso, 2013).

Furthermore, the researcher finds that more *base form of lexical verb (VV0)* is used by the females (1, 358 words) than the males (1, 077 words). Likewise, Biber et al. (1998) discover that females tend to make reference to the present tense of verbs more than males.

Having identified these POS taggers, the next step is to recognise the relationships that exist between these taggers. Hence, the researcher looks at these POS taggers by grouping them in the sequence of *Personal Pronouns*, *Lexical Verbs*, *Interjection*, *General Adverb*, and one particular form of the copular verb *be,* i.e., *am* at last.

Personal Pronouns

According to Argamon et al. (2003: 326), pronouns such as "*I, you, she, her, their, myself, yourself and herself* are all strong female indicators". Specifically, they also declare that the usage of *I, you* and *she* is much more significant in female texts than in male texts. In the following sections, the researcher elaborates gender differences in the usage of *third person singular personal pronoun* (i.e. *he, she, him,* and *her*) and *first person singular subjective personal pronoun* respectively.

1. Third Person Singular Personal Pronoun

As shown in Table 4.6, both the females and males use the *third person singular subjective personal pronoun* (i.e. *he* and *she*) twice as many as the *third person singular objective personal pronoun* (i.e. *him* and *her*). Under further scrutiny, the females share a much higher usage of the third person singular subjective personal pronouns, namely, *he* (F: 166 times, 0.40%; M: 94 times, 0.24%) and *she* (F: 121 times, 0.29%; M: 19 times, 0.05%) than the third person singular objective personal pronouns, namely, *him* (F: 67 times, 0.16%; M: 23 times, 0.06%) and *her* (F: 50 times, 0.12%; M: 15 times, 0.04%). Although Argamon et al. (2003) demonstrates that the males use more male third pronouns (i.e. *he, him*) the females, the present study affirms that it is the females rather than the males that use more male third pronouns, even more than the usage of the female third pronouns (i.e. *she, her*). However, this is inconsistent with Kaur's (2009) finding that the Malaysian girls write more about their female friends which lead

to an overwhelmingly higher usage of *she* (G: 929 times, B: 160 times) than *he* (G: 508 times, B: 559 times) in the girls' writing texts.

2. First Person Singular Subjective Personal Pronoun (PPIS1)

Examinations of previous studies suggest that first-person singular pronouns are combined with "perceptual or cognitive" verbs (e.g., "I guess") in women's writing (Mulac and Lundell, 1994; Palander-Collin, 1999; Argamon, 2003) and speech (Hartman, 1976; Poole, 1979; Holmes, 1990; Preisler, 1986; Rayson, Leech and Hodges, 1997) in order to reveal their preference for hedging. In particular, the use of the phase *I think* (F: 13 times, 0.03%; M: 6 times, 0.02%) which is also a female preference (Fahy, 2002; Schmid, 2003) mirrors the uncertainty in the female thoughts as shown in Concordance 4.33.

Concordance 4.33 Concordance line (females) of *I think* (6 of 13 lines)

N			
1	my bf is sick with indication that	*I think*	if he come ED HKL he sure be
2	std by Daniel. It's hilarious.	*I think*	my head were gonna explode from
3	Ended the party around 1 a.m.	*I think*	. I sure had a good night sleep.
3	reading Bridge of Signs which	*I think*	might endure to become the best
4	a Pixie machine, 200 capsules (	*I think*	) and a presentation box together
5	led me to stop typing as	*I think*	maybe it's not cool to share my
6	about something. Sometimes	*I think*	' Should I blog and share the stories

The other seven concordance lines show the women's confidence about their thoughts whilst the concordance lines in the male sub-corpus are equivalently divided to

express uncertainty (3 occurrences) and confidence (3 occurrences). This is consistent with Skoglund's (2009) findings that the males use hedges to show confidence and uncertainty whereas the females use it to show confidence of a higher frequency.

An interesting discovery lies in the hedges in male sub-corpus. They also use *I guess* three times (0.01%) which is balanced with its frequency in the female sub-corpus (4 times, 0.01%). Still as a support to previous studies, the females are identified to use more hedges than the male equivalents (Coates, 1993; Lakoff, 2004; Nowson, 2006; Yale, 2007; Amir et al., 2012). However, in Schmid's (2003) and Skoglund's (2009) studies, they assert that the males use the phases such as *perhaps, I guess* and *sort of* more frequently than the females, and this contradicts the present study. Overall, all the five terms (i.e. *I, I think, I mean, I guess,* and *I say*) in the female sub-corpus are encompassed in the male sub-corpus.

Lexical Verb

According to Biber et al. (1998), women organise their writings with more present-tense verbs. It is also indicated by Newman et al. (2008) that females use more verbs than the males. In the present study, two POS taggers, namely, *past form of lexical verb* and *base form of lexical verb* are grouped under this category. More in-depth findings are discussed in the following section.

1. Base Form of Lexical Verb (VV0)

Generally speaking, the females and males share a great similarity in word choices with seven types are the same including opinion-related words, e.g. *know* (F: 45 times, 0.11%; M: 38 times, 0.10%), and *think* (F: 29 times, 0.07%; M: 28 times, 0.07%); demand-related, e.g. *need* (F: 42 times, 0.10%; M: 23 times, 0.06%), *want* (F: 37 times, 0.09%; M: 26 times, 0.07%); gain-related, e.g. *get*[16] (F: 35 times, 0.08%; M: 19 times, 0.05%); and emotion-related, e.g. *thank you* (F: 26 times, 0.06%; M: 27 times, 0.07%), and *like* (F: 21 times, 0.05%; M: 14 times, 0.04%).

Three distinct verbs in the female sub-corpus are *love* (40 times, 0.10%), *let* (21 times, 0.05%) and *guess* (18 times, 0.04%). The high frequency of the word *love* refers to the fact that females record more about their feelings toward love than the males (Arici, 2009). As for the word *guess* (18 times, 0.04%), eight of eighteen concordance lines refer to the phase *I guess*, which restates female preponderance with first person singular subjective personal pronoun (*I*).

2. Past Tense of Lexical Verb (VVD)

In the female sub-corpus, the past tense of lexical verbs is also significantly used. Again, the two genders share the usage of certain verbs, such as *said* (F: 45 times,

[16] The over use of the word '*get'* probably refer to the thirteenth most significant semantic domain, namely, *Getting and possession (A9+)*, in the female sub-corpus.

0.11%; M: 18 times, 0.05%), *went* (F: 37 times, 0.09%; M: 15 times, 0.04%), *got* (F: 36 times, 0.09%; M: 29 times, 0.07%), *thought* (F: 25 times, 0.06%; M: 15 times, 0.04%), *decided* (F: 23 times, 0.06%; M: 13 times, 0.03%), *made* (F: 21 times, 0.05%; M: 16 times, 0.04%) and *took* (F: 16 times, 0.04%; M: 20 times, 0.05%). Specifically, the present and past tenses of the verb *think* are significantly used in both sub-corpora.

Interjection

In general, the females (195 terms, 0.47%) use interjections twice as many as the males (89 terms, 0.23%). The females and males use the word *yes* (F: 22 times, 0.05%; M: 17 times, 0.04%) most significantly in the sub-corpus. Other words including *haha* (F: 18 times, 0.04%; M: 5 times, 0.01%) to replace a happy smiley, *oh* (F: 18 times, 0.04%; M: 7 times, 0.02%) to express realisation, *yeah* (F: 12 times, 0.03%; M: 8 times, 0.02%), *hello* (F: 11 times, 0.03%; M: 5 times, 0.01%), *no*[17] (F: 10 times, 0.02%; M: 6 times, 0.02%) and *ah* (F: 9 times, 0.02%; M: 3 times, 0.01%) to feel sorry and pity for someone are also noticed in both sub-corpora with a higher frequency in the female sub-corpus.

[17] This is consistent with Schler et al. (2006) who found that female bloggers use more assent and negation words than the male bloggers as the word *no* occurred with frequencies of ten times (0.02%) in the female sub-corpus and six times (0.02%) in the male sub-corpus. Also identified by Koppel et al. (2002) that females tend to use more negation which characterised the female style.

Concordance 4.34 Concordance line (females) of alhamdulillah (all 9 lines)

N			
1	3. After 2 early assessments,	*alhamdulillah*	I pass! Probably not with flyin
2	saya terima sepanjang 2012	*alhamdulillah*	for them all. I need to plan fo
3	to grow up with. Mother of 2	*alhamdulillah*	.. I am now a mother of 2 girls
4	in the OT was a good one and	*alhamdulillah*	, everything went smoothly. And
5	loud or get cranky at night..	*alhamdulillah*	.. Thank you Allah.. Well will
6	“Wah!” He just loves it	*alhamdulillah*	. Bronchial Thermoplasty
7	ry best to start a collection.	*alhamdulillah*	, orders keep pouring in and I’
8	a car buy a house married haj	*alhamdulillah*	.. mission accomplished (accept
9	uced its degree of protrusion	*alhamdulillah*	. I will show you guys the results

One distinguished religious word in the sub-corpus is *alhamdulillah* (F: 9 times, 0.02%; M: twice, 0.01%) which is used in Arabic to show gratitude to *Allah* (Concordance 4.34).

Additionally, the females use the exclamatory emotive interjection *wow* (8 times, 0.02%) to express the feeling of amazement, while one male use it once in the male sub-corpus. This is in accordance with Bednarek’s (2009) findings that the female characters (1009 occurrence) use much more exclamatory emotive interjections than their male counterparts (239 occurrence) in TV series. However, Ali and Aslam (2012) illustrated that the Pakistani men use more interjections in code mixed SMS than the women.

General Adverb (RR)

As previous studies notified that women are the prolific uses of the intensive

adverbs such as *just, so, really, very* and *vastly* (Jespersen, 1922; Key, 1975; Lakoff, 1975; Aries, 1998; Amir et al., 2012), the present study also replicates the results with higher frequencies of intensifiers such as *so* (F: 127 times, 0.31%; M: 76 times, 0.19%), *really* (F: 79 times, 0.19%; M: 67 times, 0.17%), *only* (F: 62 times, 0.15%; M: 42 times, 0.11%), and *always* (F: 45 times, 0.11%; M: 37 times, 0.09%) which are in accordance with previous research studies (Fahy, 2002; Sharp, 2012). The most significantly occurring word is *just* in both sub-corpora with frequencies of 148 times (0.36%) in the female sub-corpus, and 111 times (0.28%) in the male sub-corpus.

Likewise, the females are noticed to use more negation words than the males with a high frequency of the word *never* (39 times, 0.09%) in their sub-corpus, while the males use the word *never* 28 times (0.07%) which is consistent with the results of Newman et al. (2008) that women use the negation word *never* more frequently than the male equivalents. However, Fahy (2002) indicates that males tend to use the intensifier, i.e. *never*, more than the females.

Am (VBM)

According to McGinnis, Goodstein-Stolzenberg, and Saliani (2007), abbreviations are significantly used in CMC contexts. Therefore, the frequencies of *am* (83 times, 0.20%) and *'m* (82 times, 0.20%) are approximate in the female sub-corpus. Contradicting to the females, the males use the abbreviated form of am, i.e. *'m* (49 times,

0.12%) less frequently. This finding supports previous research where evidence of gender differences is identified in the use of abbreviations (Herring and Zelenkauskaite, 2009; Halmetoja, 2013). Concerning the *am*-phase, *am about to* occurs once in both sub-corpora with other two phases, namely, *am back* (3 times, 0.01%) and *'m back* (once) solely used by the male bloggers.

4.4.2 Key Parts of Speech in the Male Weblogs Sub-corpus

Table 4.7 Frequency Lists for Key Parts of Speech in Male Weblogs

Rank	**Tag**	**Key Part of Speech**	**Frequency**	**%**
1	NN1	singular common noun	5530	14.10
2	JJ	general adjective	2677	6.83
3	II	general preposition	2141	5.46
4	AT	article	1937	4.94
5	NN2	plural common noun	1567	4.00
6	RR	general adverb	1435	3.66
7	VVI	infinitive	1317	3.36
8	CC	coordinating conjunction	1263	3.22
9	VV0	base form of lexical verb	1077	2.75
10	PPIS1	1st person sing. subjective personal pronoun (*I*)	1041	2.65

Table 4.7 illustrates the ten most significantly occurring POS in the male sub-corpus with the same types in the female sub-corpus but with different tokens. As for the male indicators, article (AT: 1937 terms, 4.94%), and general preposition (II: 2141 terms, 5.46%) are of higher frequency in the male sub-corpus. They also use more general adjective (JJ: 2677 terms, 6.83%) than their female counterparts. Concerning the other POS, females use them more frequently than males.

Table 4.8 Key POS in Male Weblogs Compared to Female Weblogs

Rank	**Tag**	**Part-of-speech**	**Male weblogs**		**Female weblogs**		**LL value**
			Freq	***%***	***Freq***	***%***	
1	IO	of (as prepositions)	752	1.92	572	1.38	36.21
2	AT	article	1937	4.94	1690	4.06	34.39
3	VBZ	is	738	1.88	587	1.41	27.21
4	FO	formula	161	0.41	87	0.21	26.96
5	JJ	general adjective	2677	6.83	2496	6.00	21.35
6	VBR	are	248	0.63	173	0.42	18.18
7	NP1	singular proper noun	1009	2.57	891	2.14	15.86
8	II33	the preposition at the third place of a three-word phrase (e.g. '*as*' in '*is defined as*')	10	0.03	0	0.00	14.46
9	II31	the preposition at the first place of a three-word phrase (e.g. '*in*' in '*in front of*')	29	0.07	9	0.02	12.28
10	II	general preposition	2141	5.46	2038	4.90	12.14
11	II22	the preposition at the second place of a two-word phrase (e.g. '*of*' in '*one of*')	8	0.02	0	0.00	11.57

According to Table 4.8, significant gender differences emerge in the use of prepositions when the males are compared to the females. There are five POS taggers (IO, II33, II31, II, and *II22*) in total and the overuse of prepositions is in accordance with the findings of related work (Newman et al., 2008; Ali and Aslam, 2012; Schler et al., 2006; Rangel and Rosso, 2013).

While the greater usage of the most salient male indicator, namely, *preposition* (Biber et al., 1998; Koppel et al., 2002) is clear, closer scrutiny verifies that the use of *of*

(M:751 times) and *of-related phrases* (e.g. *of the day*) is of the most striking gender difference when males are compared to females.

Another male indicator, the article, also bears a higher usage in the male sub-corpus than that of the female sub-corpus, which is in accordance with previous research (Newman et al., 2008; Flekova and Gurevych, 2013; Yu, 2013; Rangel and Rosso, 2013).

As for general adjectives, the present study consents with Rangel and Rosso's (2013) results that males use them more than females but contradicts other findings where females use more adjectives than their male counterparts (Ali and Aslam, 2012; Yu, 2013). Although Biber et al. (1998) claim that males use more nouns than females, Yu (2013) clarifies that it is the females rather than males who use more nouns. While in this study, the researcher confirms the former statement and shares the result with Argamon et al. (2003) that males use more proper nouns than their female counterparts.

Likewise, the researcher conducts her analysis in six categories, namely, *Prepositions*, *Article*, *General Adjective*, *Singular Proper Noun*, Copular verb *be*, and *Formula*.

Prepositions

While previous studies labels preposition as a male indicator (Biber et al., 1998; Koppel et al., 2002; Schler et al., 2006; Rangel and Rosso, 2013), the POS of the present study also encompasses a large proportion of prepositions in the male sub-corpus.

1. *of* (as prepositions) (IO)

The most significant gender difference lies in the *of*-phrase which is of higher frequency in the male sub-corpus (752 occurrences, 1.92%) than that of the female sub-corpus (572 occurrences, 1.38%). Although Argamon et al. (2003) indicate that males tend to use more *of*-phrase, the insufficient concordance lines in the present study fail to prove this statement. There is only one *of*-phrase tagged in the male sub-corpus, namely, *of the day* (once) which also occurs once in the female sub-corpus. Another *of*-phase is *of a time* appears once in the female sub-corpus. In general, the males (751 times, 1.91%) use the word *of* much more than the females (570 times, 1.37%) in the sub-corpus respectively.

2. General preposition (II)

As for the choice of exact prepositions, the males tend to use certain prepositions more than the female counterparts including *in* (M: 518 times, 1.32%; F: 397 times, 0.95%), *on* (M: 196 times, 0.50%; F: 186 times, 0.45%), *by* (M: 106 times, 0.27%; F: 97

times, 0.23%), *about* (M: 94 times, 0.24%; F: 88 times, 0.21%), and *like* (M: 79 times, 0.20%; F: 73 times, 0.18%). On the contrary, prepositions including *to* (M: 325 times, 0.83%; F: 345 times, 0.83%), *from* (M: 149 times, 0.38%; F: 155 times, 0.37%), *at* (M: 141 times, 0.36%; F: 176 times, 0.42%), and *after* (M: 44 times, 0.11%; F: 53 times, 0.11%) occur more frequently in the female sub-corpus rather than that of the male sub-corpus. Two distinct words in the corpora are *as* (52 times, 0.14%) in the male sub-corpus and *before* (48 times, 0.12%) in the female sub-corpus. In general, the males (2141 occurrences, 5.46%) use more prepositions than the females (2038 occurrences, 4.90%) which supports previous researches (Koppel et al., 2002; Pennebaker, Mehl, and Niederhoffer, 2003; Schler et al., 2006; Newman et al., 2008; Morgan, Banks, and Boals, 2011; Rangel and Rosso, 2013).

Article (AT)

The present study corroborates past researches (Nowson, 2006; Schler et al., 2006; Newman et al., 2008; Flekova and Gurevych, 2013; Yu, 2013) in that the males are prolific users of articles when compared with the females. The words *the* and *no* occur most frequently with frequencies of 1859 times (4.74%) and 57 times (0.15%) in the male sub-corpus and 1622 times (3.90%) and 49 times (0.12%) in the female sub-corpus. Another two phases shared by the two genders are *the bill* (M: 3 times, 0.01%; F: twice) and *the sun* (M: twice, 0.01%; F: once) (Concordance 4.35).

Concordance 4.35 Concordance line (males and females) of *the sun* (all 3 lines)

N			
1	clear, you would be waiting till	*the sun*	comes down! Thanks to Hanoi kids,
2	the very 1st day of spring when	*the sun*	is exactly at the celestial longitude of
3	to catch this beautiful horizon.	*the sun*	was considerably merciful on us near

As implied in Concordance 4.35, the first two lines are extracted from the male sub-corpus while the last line is an excerpt in the female sub-corpus. All of the three concordance lines refer to the scenery during the journey but the differences lie in the stories behind the view. The first line makes reference to the travel in Hanoi and the second line explains the original story of the Chinese New Year, whilst the last line records the journey in Bukit Jugra which is a historical town and a former royal town in the state of Selangor. Again, this corroborates with previous findings that Malaysian males in the present study prefers to travel abroad and females favour domestic journey. Moreover, the males tend to elaborate the contexts under a historic basis.

General Adjective (JJ)

A retrospect of previous work on gender difference in adjectives identifies that females tend to use more empty adjectives than males (Jespersen, 1922; Lakoff, 1975; Quicke, 2012; Cholifah, Heriyanto and Citraresmana, 2013), even in the Malaysian blogosphere (Amir et al., 2012). However, the use of general adjectives are controversial as some researchers indicate that females use adjectives more frequently than males (Ali and Aslam, 2012; Yu, 2013) whereas others imply that males are prolific

users of adjectives rather than females (Rangel and Rosso, 2013). The present study upholds the latter finding that the males (2677 occurrences, 6.83%) tend to use general adjectives slightly more than the females (2496 occurrences, 6.00%). The words *good* (M: 67 times, 0.17%; F: 64 times, 0.15%), *new* (M: 42 times, 0.11%; F: 53 times, 0.13%), and *happy* (M: 40 times, 0.10%; F: 50 times, 0.12%) are the most significantly used terms in both the sub-corpora. Another two adjectives shared by the sub-corpora are *big* (M: 35 times, 0.09%; F: 36 times, 0.09%), and *old* (M: 24 times, 0.06%; F: 35 times, 0.08%) with higher frequencies in the female sub-corpus compared to that of the male sub-corpus.

Furthermore, the males use the word *Chinese* to a great extent with a frequency of 29 times (0.07%). In particular, fifteen of the twenty-nine concordance lines refer to Chinese New Year, four concordance lines refer to Chinese restaurants and food, and two concordance lines refer to Chinese tea respectively. Therefore, the phase *New Year* is of significant usage in the male sub-corpus (27 times, 0.07%), which is also evident in the female sub-corpus (23 times, 0.06%). However, the foci of the two sub-corpora, is different. The males make reference to the celebration of Chinese New Year (13 occurrences) and seven concordance lines refer to the celebration of a new year with only two concordance lines relating to the new year's resolution, whilst the females write more about their new year's resolutions (6 occurrences) with seven concordance

lines relating to the Chinese New Year and seven concordance lines refer to the celebration of a new year.

One interesting result is that the males (23 times, 0.06%) use the word *beautiful* more than the females (18 times, 0.04%) who are stereotyped to talk more about appearances.

Concordance 4.36 Concordance line (males) of *beautiful* (21 of 23 lines)

N			
1	this dressing up is to make us	*beautiful*	. I know, little one, we are
2	now, little one, we are already	*beautiful*	is exactly at the celestial
3	to catch this beautiful horizon.	*beautiful*	. However, for some strange
4	and makeup make the	*beautiful*	. What is beautiful? Good
5	question, little one. To be	*beautiful*	is to make your physical
6	books, the pictures of	*beautiful*	women in the past are very fat!
7	thin! Where once being fat is	*beautiful*	, now it is beautiful to be skinny
8	being fat is beautiful, now it is	*beautiful*	to be skinny. I know, it is very
9	hopefully not for too long.	*Beautiful*	or cute? What would you prefer
10	before his is the most stunning,	*beautiful*	, sexy woman he has ever seen in
11	a person think, feel and act. A	*beautiful*	soul is full of love, joy, peace
12	believes that if the inner soul is	*beautiful*	, then the outer appearance will
13	may you grow to up to be a	*beautiful*	soul. May you always have
14	I was blessed enough to have	*beautiful*	people around me. If they had n't
15	mystery that has shrouded this	*beautiful*	country from the rest of the world
16	presentation of the varied and	*beautiful*	Spilok Forest, which the Forest
17	Cruising along the river with	*beautiful*	karst formations on both sides
18	Spirit. I believe that it is a	*beautiful*	church. The painting at the altar
19	city of Yogyakarta. The	*beautiful*	Dieng Plateau, one of the highlig
20	but again, we have to give this	*beautiful*	island a miss caused of the ferry
21	Malacca Riverbank because of	*beautiful*	night view, and this was the hote

As shown in Concordance 4.36, lines 1 to 10 refer to appearance, lines 11 to 14 refer to soul and lines 15 to 21 refer to scenery. Whilst eleven of the eighteen concordance lines in the female sub-corpus make references to scenery, two lines are related to the bloggers' feelings, one line refers to photos and only two lines make references to the appearance of the bloggers' children.

Singular Proper Noun (NP1)

Although Yu (2013) states that females use more nouns than males, the present study supports previous studies that males use nouns more frequently than females (Biber et al, 1998; Argamon et al., 2003). Regarding singular proper noun (NP1), the most frequently occurring word is *Malaysia* - 28 times (0.07%) in the male sub-corpus and 17 times (0.04%) in the female sub-corpus respectively. Other place-related words in the male sub-corpus are *Hanoi* (20 times, 0.05%), *Myanmar* (12 times, 0.03%), *Seoul* (10 times, 0.03%), *New York* (9 times, 0.02%), *Singapore* (8 times, 0.02%) and *Gangnam* (7 times, 0.02%) which reconfirm the point that males write more about the place when compared with the females ((Mulac et al., 1990; Johnstone, 1993; Mulac and Lundell, 1994). To be more specific, males write more about international places, whereas females concentrate on domestic areas such as *KL* (11 times, 0.03%), *Bukit Jugra* (8 times, 0.02%), and *Penang* (7 times, 0.02%) which is in accordance with the previous section of the present study.

Both the males and the females mention about personal names, however, the foci is different. While the three names in the female sub-corpus, namely, *Henry* (12 times, 0.03%), *Peter* (8 times, 0.02%) and *Ethan* (7 times, 0.02%) are all about the bloggers' travelling companies, the name, *Altantuya* (13 times, 0.03%) is much more meaningful. It is about a mysterious murder which happened on October 18, 2006 when a Mongolian lady was murdered by C-4 explosives. Some politicians were rumoured to be involved in this scandal but the story ended with a controversial verdict. Therefore, the public show some resentment to the government which is unfolded from the concordance line of the word *PM* (Prime Minister) (Concordance 4.37). As other scholars (Amir et al., 2012; Jespersen, 1922) implied, males tend to favour public issues rather than personal topics which are preferred by females.

Concordance 4.37 Concordance line (males) of *PM* (4 of 10 lines)

N			
1	No thanks, of course, to Our	*PM*	, who simply was n't inspired enough
2	will have to wait until Our Dear	*PM*	is inspired enough, but constitution
3	trail now leads suspiciously to	*PM*	and to the circle of people close to
4	this is a private matter that the	*PM*	must sort out himself, having made

Copular Verb (be)

Two parts of speech are under this category. One is the verb *is*, the other one is the verb *are*. In general, the frequencies of *is* and *is*-phase are higher than *are* and *are*-phase, but the males use the verb *is* (M: 602 times, 1.53%; F: 449 times, 1.08%) and *are* (M: 236 times, 0.60%; F: 162 times, 0.39%) more significantly than the females.

1. Is (VBZ)

The two most frequently occurring words, namely, *is* and its abbreviation, *'s*, are shared by the two sub-corpus with frequencies of 602 times (1.53%) and 128 times (0.33%) in the male sub-corpus and 449 times (1.08%) and 128 times (0.31%) in the female sub-corpus. Due to the low frequency of the related phases, no significant gender differences are located from the sub-corpora.

2. Are (VBR)

The results of the comparison between the two sub-corpora about the usage of the verb show that the males and females use the word *are* (M: 236 times, 0.60%; F: 162 times, 0.39%) and *'re* (M: 236 times, 0.60%; F: 162 times, 0.39%) most frequently. However, there is only one phase that is tagged in the male sub-corpus with no concordance lines in the female sub-corpus which makes it irrelevant to analyse the gender differences between the two sub-corpora.

Formula (FO)

In the present study, the males (16 terms, 0.41%) use the formula semantic domain twice as many as the females (87 terms, 0.21%). The most significantly occurring type in both sub-corpora are symbols such as '=' (8 times, 0.02%), in the male sub-corpus and '+' (9 times, 0.02%), in the female sub-corpus. Specifically, the occurrences of the types are in accordance with the relevant domains discussed in the previous section.

4.5 Discussions of the Findings

In terms of the findings presented in previous sections, significant gender differences are noticeable from the analysis which encourages the researcher to seek the factors that lead to those gender differences. Hasan and Khammat (2011) analyzed the socio-cultural factors with *difference theory* and *dominance theory* cited in Nemati and Bayer (2007: 2). Hasan and Khammat's (2011) research shares some similarities with the present study in two aspects:

1) Both of the two studies investigated the linguistic features in ESL community Hansan and Khammat (2011) examined the gender differences in the usage of sentence type and sentence modifiers by Iraqi ESL learners in a university while the present study examines the gender differences in semantic domains and parts of speech in Malaysian ESL adults.

2) Both of the two studies conduct a content analysis of asynchronous CMC as Hansan and Khammat (2011) investigated the content of emails and the present study investigates weblogs.

Accordingly, the researcher adopts the two theories as the theoretical bases to explain the factors that lead to the gender differences in the present study.

From the cultural perspective, men and women are from different sub-cultures and the socialization processes they experienceearly on are also different. As a cultural product, women speak a language of connection and intimacy. Men speak a language of status and independence (Tannen, 1990). In short, *culture* governs how we think and feel, how we behave and how we live, and it is born largely of socialization. *Gender* roles provide a primary way to classify social life within a culture. In the female sub-culture, women are taught to be tender, passive, non-competitive and considerate, thus when grown up, they tend to be more tender, non-competitive, considerate, but not as good at solving problem, less likely to assume leadership and unwilling to take risks; whereas in the male sub-culture, men are expected to be independent, direct, aggressive and risk-taking, thus they are inclined to be confident, direct, aggressive, good at solving problem, more likely to assume leadership and willing to take risks.

These different cultural expectations from others affect their linguistic characteristics. For example, from childhood, girls are taught to speak and act “like a lady”. A girl would surely be scolded when a taboo word came out from her mouth; however, chances would be much less for a boy to be scolded when he did the same thing because “boys are boys”. Thus women tend to use no or less taboo words in daily life as well as online to keep their image of a “lady”.

With regard to the third research question, the social context of Malaysia facilitates gender differences between the females and males in the present study. Family related domains such as *Food (F1), health (B3, B1, B2-), Kinship (S4)* and *residence (H2)* are predominant in the female sub-corpus. The males on the other hand focus on work which can represent their social status. Statistics from *The 2013 Global Gender Gap Report*[18] (p. 268) have shown that 79% of Malaysian males are in labour force compared to 46% of Malaysian females among the population of 28.86 million. This point indicates that the males in Malaysia are the primary workforce. In other words, their dominance in society shapes their language to reflect their social roles. Evidences can be notified in the domain of *Money: Affluence (I1.1+)* and *Money and pay (I1.1).*

Regarding online communication within communities, females and males are influenced by these two factors. Females seek association by blogging about personal affairs such as the domain of *People: Female (S2.1)* and use a large number of pronouns to refer to their stories. On the contrary, the male communication is more impersonal and they are confident to reveal their personalities by domains of *Participating (S1.1.3+)* and *Competition (S7.3).* This is the cultural offspring to create femininity in women's childhood and the social process to endow women less power than men (Hansan and Khammat, 2011). As a result, the female language is keen on acceptance whereas the male language seeks social status.

[18] Source: http://www3.weforum.org/docs/WEF_GenderGap_Report_2013.pdf

CHAPTER 5

CONCLUSION

5.1 Introduction

In this chapter, the researcher concludes her research on semantic domain differences between female and male weblogs in an ESL context. The study uses the corpus-driven method to examine the extent to which females and males in this particular cultural setting, i.e. Malaysia, use semantic domains and parts of speech differently. In the following sections of this chapter, the researcher presents a summary of significant findings (Section 5.2), and implications of the study (Section 5.3), followed by reflections on the research which also deals with limitations for the study (Section 5.4) and suggestions for future research (Section 5.5).

5.2 Significant Semantic (i.e. semantic domains) and Stylistic Features (i.e. POS)

Overall, the present study tags the key semantic domains and key parts of speech in female and male sub-corpora which are analysed in Sections 4.3 and 4.4. By referring to related research about gendered semantic and stylistic features, further discussions are presented in the following sections.

5.2.1 The Female Content

The present study consolidates Herring's (1993) statement that families and personal affairs are women's preferred online topics. Significant semantic features in the

present study refer to *people and relationships*, *body and health*, and *food* (which is the most significant semantic domain in the present study).

Malaysian females record their daily life in the category of personal weblogs. Key semantic domains make references to interpersonal communication with people in the community they involve in, such as *Kin (S4)*, and *People: Female (S2.1)*, as well as their current concerns about their family members, such as *Medicines and Medical treatment (B3)*, *Disease (B2-)*, and *Food (F1).*

All the significant semantic features indicate that the Malaysian females, like other women in other linguistic, cultural and social backgrounds, share the same written style which strengthens the involvement to the community and the cooperation with other members.

The females in the present study use significant POS compared to their male counterparts, such as third person singular personal pronouns (*he, she, him, and her*), interjection, past tense and base form of lexical verb. The words, *I* and *am*, are also identified of a higher frequency in the female sub-corpus than that of the male sub-corpus.

An unexpected result is that the usage of *general adverb*, which is identified as a male indicator in relevant studies, is more significant in the female sub-corpus in the present study.

5.2.2 The Male Content

Herring (1993) illustrates that males are keen on talking about politics and providing information online. In the present study, Malaysian males write blogs about public issues with the *G3* tag *Warfare, defence and the army; weapons*.

As *Quantity* and *Place* are regarded as male predilection (Mulac, Studley, and Blau, 1990; Mulac and Lundell, 1994), they are also the significant semantic features in the present study. In the category of *Geography and Location*, references are made to *Geography names (Z2)*, *Geographical terms (W3)*, and *Places (M7)* whereas the *N5+* tag *Quantities: many/much* marks the quantifiers used by the Malaysian males.

Other features are related to the male work, how they make a living (*I1.1: Money and pay*), how they participate in business activities (*S1.1.3+: Participating*) and compete with colleagues (*S7.3: Competition*). Their styles of communication with the people from the community are more impersonal rather than interpersonal, which is a convention among females. They seek information from the communication and provide information in their writings.

Interesting finding points towards leisure activities as the males in the present study blog about the *Music and related activities (K2)* domain which is not so pervasive in previous studies.

Prepositions dominate the key POS in the male sub-corpus with the preposition, *of*, producing the most significant gender difference. Moreover, articles and singular proper nouns are more frequently used in the male sub-corpus. Two verbs, namely, *is* and *are*, are also of males' preference.

Surprisingly, the Malaysian males in the present study use more general adjectives, in their blogs. They also use the word *beautiful* referring to appearance.

5.3 Implications of the Thesis

The blogosphere is a powerful arena for public social networks. The potential of blogs as a strategic communication tool stems from recent data that show reading blogs is becoming popular among the general population (Segev, Villar, and Fiske, 2012). The importance of studying blogs as a social tool has been widely recognized (Kent, 2008; Porter, Sweetser, and Chung, 2009; Smith, 2011; Xifra and Huertas, 2008). Researchers argue that blogs possess several characteristics that make them a powerful strategic communication tool, but of primary importance is the fact that blogs form communities

of coherent groups of individuals and professionals with shared common interests (Kent, 2008).

Taking gender into account, the present study seeks to facilitate online communication with more effectiveness and politeness between women and men which is set as the practical implication of the research. In an ESL context, this study also aims to instruct communication between females and males in Malaysia, a multi-ethnic and multi-racial country.

In contrast to more traditional environments, technology offers greater opportunities for interactivity and learner control (Kozma, 1991; Rodzvilla, 2002). There are more educators and language teachers using the Internet in language teaching as well (Godwin-Jones, 2003; Lord and Lomicka, 2004). Many computer applications, especially asynchronous computer-mediated communication, promote interactive learning (Arnold and Ducate, 2006). With the booming growth of technology, blogs have become another learning platform for language teaching (Richardson, 2005). Furthermore, a blog is interactive in the sense that readers can respond with comments in just a few steps (Rodzvilla, 2002). Therefore, there have been increasing numbers of people using blogs in education (Richardson, 2005). Blogs are well suited to serve as online journals for users (Godwin-Jones, 2003; Richardson, 2005). In terms of language teaching and learning, blogs are also supportive in the sense that "language learners

could use a personal blog linked to a course as an electronic portfolio, showing development over time" (Godwin-Jones, 2003: 13).

Pedagogically, in the present study, studying blogs in an ESL context could be used to monitor and assess ESL learners as well as to encourage interaction among students and between teachers and students. A teacher can draw up a more appropriate teaching plan for male and female learners regarding individual differences in order to promote more effective ESL learning.

5.4 Limitations

The methodology, by virtue of being able to make statistical comparisons between corpora, is able to generate list of words from the two data sets of female and male weblogs in one culture for comparative purposes. However, like any other methodology, the method used in this research has a number of limitations.

The duration of the whole schedule pertaining to the dissertation limits the period of data collection. Therefore, the researcher solely focuses on one particular period of time (September 2012 to April 2013) which might not represent the phenomenon of the whole Malaysian blogosphere.

Another variable that is beyond the researcher's control is the choices of the bloggers. The participants of the present study are not further grouped into races or religions due identification issues. As a result, this small group of people need not necessarily reflect the blogging behaviours of all Malaysian adult bloggers.

5.5 Suggestions for Further Research

To compensate for the limitations, it is recommended that future studies be conducted with an increased number of participants or in an organized group which can level off the influence of this kind of inaccuracy, e.g. a longitudinal study of selected bloggers could be conducted.

It would also be interesting to carry out a study to make a comparison between the online blog language with real life spoken and written language or a comparison between an ESL community and an ENL community in order to investigate whether blog language has certain special linguistic features among different spheres or settings.

REFERENCES

Abd. Samad, A., Hassan, F., Mukundan, J., Kamarudin, G., Syd Abd. Rahman, S. Z., Md. Rashid, J., & Vethamani, M. E. (2002). *The English of Malaysian school students (EMAS) corpus*. Serdang, Malaysia: Universiti Putra Malaysia.

Adhami, A., Khademian, T., Almasi, H., & Rafiei, L. (2012). An analysis on editorial content of two Iranian newspapers, regarding the parliamentary election campaigns for the Ninth Islamic Consultative Assembly. *Journal of Basic Applied Sciences Research, 2*(20), 10706-10712.

Afida Mohamad Ali (2005). *A comparative study of business: Journalism style in 'Malaysian business and management today': A functional textual approach.* (Unpublished doctoral dissertation). The University of Lancaster, Lancashire, United Kingdom.

Ali, I., & Aslam, T. M. (2012). Frequency of learned words of English as a marker of gender identity in SMS language in Pakistan. *Journal of Elementary Education, 22* (2), 45-55.

Amir, Z., Abidin, H., Darus, S., & Ismail, K. (2012). Gender differences in the language use of Malaysian teen bloggers. *GEMA Online Journal of Language Studies, 12*(1), 105-124.

Andersson, I., Gauding, J., Graca, A., Holm, K., Öhlin, L., Marklund, U., & Ericsson, A. (2011). Productive vocabulary size development in children aged 18-24 Months-Gender Difference. *Fonetic, 51*, 109-112.

Archer, D., Wilson, A., & Rayson, P. (2002). *Introduction to the USAS category system*. Retrieved from http://ucrel.lancs.ac.uk/usas/usas guide.pdf

Argamon, S., Koppel, M., Fine, J., & Shimoni, A. (2003). Gender, genre, and writing style in formal written texts. *Text, 23*(3), 321-346.

Arici, A. F. (2009). Preferred writing Topics of Pre-service Teachers. *European Journal of Educational Studies, 1*(2), 95-99.

Aries, E. (1998). Gender differences in interaction: A reexamination. In D. J. Canary & K. Dindia (Eds.), *Sex differences and similarities in communication* (pp. 65-81). Mahwah, NJ: Erlbaum.

Armstrong, K., & Retterer, O. (2008). Blogging as L2 writing: A case study. *The Journal of Association for the Advancement of Computing in Education, 16*(3), 233-251.

Arnold, N., & Ducate, L. (2006). Future foreign language teachers' social and cognitive collaboration in an online environment. *Language Learning & Technology, 10*(1), 42-66.

Arthur, J. B. (2012). *Framing child neglect: A content analysis of national and regional U.S. newspaper coverage*. (Unpublished master's thesis). American University, Washington, D. C., USA.

Bahiyah Dato' Haji Abdul Hamid, Mohd. Subakir Mohd. Yasin, Kesumawati Abu Bakar, Yuen, C. K., & Azhar Jalaluddin. (2008). Linguistic sexism and gender role stereotyping in Malaysian English language textbooks. *GEMA Online Journal of Language Studies, 8*(2), 45-78.

Bahrani, T., & Tam, S. S. (2012). Informal language learning setting: Technology or social interaction? *The Turkish Online Journal of Educational Technology, 11*(2), 142-149.

Bakar, N. A., Latif, H., & Ya'acob, A. (2010). ESL students' feedback on the use of blogs for language learning. *3L The Southeast Asian Journal of English Language Studies, 16*(1), 120-141.

Baker, P. (2004). Querying key words: Questions of difference, frequency and sense in key words analysis. *Journal of English Linguistics, 32*(4), 346-59.

Barjesteh, H., Vaseghi, R., Hashemi, H., Pourshahian, B., & Kalajahi, S. A. R. (2012). Gender difference in the employment of various stereotypes on Iranian English weblogs. *Language in India, 12*, 99-109.

Bausch, P., Haughey, M., & Hourian, M. (2002). *We blog: Publishing online with weblogs*. Indianapolis, IN: Wiley.

Bednarek, M. (2009, October). *Emotional practices and character identity in American popular culture.* Paper presented at the Conference of American Association for Corpus Linguistics (AACL 2009), Edmonton, Canada.

Berelson, B. (1952). *Content analysis in communication research.* New York, NY: Free Press.

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

Bimber, B (2000). Measuring the gender gap on the Internet. *Social Science Quarterly, 81*(3), 868-876.

Bidin, N. A., & Mustaffa, N. (2012). Blogosphere: How youth perceived blogs credibility. *Malaysian Journal of Communication, 28*(1), 33-54.

Blood, R. (2002). *The Weblog Handbook: Practical Advice on Creating and Maintaining Your Blog.* Cambridge, MA: Perseus Publishing.

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

Botley, S. P., Metom, L., & Dillah, D. (2007). *A corpus-based archive of learner English in Sabah/Sarawak: CALES Phase 2*. Final project report, Unit for Research, Development and Commercialisation, Universiti Teknologi MARA, Malaysia.

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

Braun, T. (2007). *Evaluations of individual scientists and research institutions: Scientometrics guidebooks series.* Budapest, Hungary: Akademiai Kiado.

Brend, R. (1975). Male-female intonation patterns in American English. In B. Thorne & N. Henley (Eds.), *Language and sex: Difference and dominance* (pp. 84-87). Rowley, MA: Newbury House Publishers.

Brennan, S. E. (1991). Conversation with and through computers. *User modelling and user-adapted interaction, 1*(1), 67-86.

Byrd, D. (1992). Preliminary results on speaker dependent variation in the TIMIT database. *Journal of the Acoustical Society of America, 92*(1), 593-596.

Byrne, M. (2005). *Workplace meetings and the silencing of women: an investigation of women and men's different communication styles and how these influence perceptions of leadership capability within Australian organisations*. (Unpublished doctoral dissertation). University of Western Sydney, New South Wales, Australia.

Cakir, I. (2006). Developing cultural awareness in foreign language teaching. *Turkish Online Journal of Distance Education, 7*(3), 154-161.

Cameron, D., McAlinden, S., & O'Leary, K. (1989). Lakoff in context: the social and linguistic functions of tag questions. In J. Coates & D. Cameron (Eds.), *Women and their speech communities* (pp. 74-93). London: Longman.

Cameron, D. (1992). *Feminism and linguistic theory*. London: Macmillan.

Cameron, D. (1996). The language-gender interface: Challenging co-optation. In V. L. Bergvall, J. M. Bing & A. F. Freed (Eds.), *Rethinking language and gender research: theory and practice* (pp. 31-53). London: Longman.

Cameron, D. (1997). Performing gender identity: Young men's talk and the construction of heterosexual masculinity. In S. Johnson & U. H. Meinhof (Eds.), *Language and masculinity* (pp. 47-64). Oxford: Blackwell.

Carlson, S. (2003, November 28). Weblogs come to the classroom. *The Chronicle of Higher Education*, pp. A33-A34.

Chang, Y. H., Chang, C. Y., & Tseng, Y. H. (2010). Trends of science education research: an automatic content analysis. *Journal of Science Education and Technology, 19*, 315-331.

Chesebro, J. W., & Bonsall, D. G. (1989). *Computer-Mediated Communication: Human Relationships in a Computerized World*. Tuscaloosa, AL: The University of Alabama Press.

Chiu, C. Y. (2008). An investigation of gender differences in EFL college writing. In M. Edwardes (Ed.), *Proceedings of the BAAL annual conference 2008* (pp. 25-26). London: Scitsiugnil Press.

Cho, S., & Huh, J. (2010). Content analysis of corporate blogs as a relationship management tool. *Corporate Communications: An International Journal, 15*(1), 30-48.

Cholifah, Y. W., Heriyanto, & Citraresmana, E. (2013). Strong expletives, empty adjectives and tag questions usage as a gender marker in a serial film "F.R.I.E.N.D.S". *International Journal of Language Learning and Applied Linguistics World, 4*(4), 403-417.

Chou, C. C. (2002, January). *A comparative content analysis of student interaction in synchronous and asynchronous learning networks.* Paper presented at the 35th Hawaii International Conference on System Sciences, Hawaii, USA.

Clavio, G., & Eagleman, A. N. (2011). Gender and sexually suggestive images in sports blogs. *Journal of Sport management, 7*, 295-304.

Coates, J. (1986). *Women, men and language*. London: Longman.

Coates, J. (1993). *Women, men and language: A sociolinguistic account of gender differences in language*. New York, NY: Longman.

Cohen-Almagor, R. (2011). Internet history. *International Journal of Technoethics, 2*(2), 45-64.

Coleman, R., & Wasike, B. (2004). Use of visual elements in public journalism newspapers during an election: A content analysis of the photographs and graphics in Campaign 2000. *Journal of Communication, 54*(3), 456-473.

Crawford, L., Pollack, J., & England, D. (2006). Uncovering the trends in project management: journal emphases over the last 10 years. *International Journal of Project management, 24,* 175-184.

Crystal, D. (2001). *Language and the Internet*. 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.

Daly, N., & Warren, P. (2001). Pitching it differently in New Zealand English: Speaker sex and intonation. *Journal of Sociolinguistics, 5,* 85-96.

Dalampan, A. E. (2006). Gender Issues in Computer-Mediated Communications. *TESL Working Paper Series, 4*(2), 59-66.

Davalos, D. B., Davalos, R. A., & Layton, H. S. (2007). Content analysis of magazine headlines: Changes over three decades? *Feminism Psychology, 17*(2), 250-258.

Davies, J. (2003). Expressions of gender: an Analysis of pupils' gendered discourse styles in small group classroom discussions. *Discourse and Society, 14*(2), 115-32.

Dibbell, J. (1993). A rape in cyberspace or how an evil clown, a Haitian trickster spirit, two wizards, and a cast of dozens turned a database into a society. *Village Voice, 21*, 36-42.

De Beauvoir, S. (1989) *The second sex*. (H. M. Parshley, Trans.). New York, NY: Vintage Books. (Original work published 1949)

Eckert, P., & McConnell-Ginet, S. M. (2003). *Language and gender.* New York, NY: Cambridge University Press.

Fahy, P. J. (2002). Use of linguistic qualifiers and intensifiers in a computer conference. *The American Journal of Distance Education, 16*(1), 5-22.

Fallon, F., Williamson, A., & Pack, M. (2011). Gender and Digital Politics. *Digital Papers 5*. Retrieved from

http://toomuchtosayformyself.files.wordpress.com/2011/07/digitalpaper-genderand digitalpolitics-2011.pdf

Fallows, D. (2005). Search engine users. *Pew Internet & American Life Project*. Retrieved from http://www.pewinternet.org/files/old-media/Files/Reports/2005/PIP_Searchengine_ users.pdf.pdf

Fausto-Sterling, A. (2000). *Sexing the body: Gender politics and the construction of sexuality*. New York, NY: Basic Books.

Feng, Y., & Karan, K. (2011). The global and local influences in the portrayal of women's roles: Content analysis of women's magazine in China. J*ournal of Media and Communication Studies, 3*(2), 33-44.

Ferdig, R., & Trammell, K. (2004). Content delivery in the "blogosphere." *The Journal of Technological Horizons in Education, 31*(7), 12-20.

Festa, P. (2003). Blogging comes to Harvard. *CNET News.com*. Retrieved from http://news.com.com/2008–1082–985714.html

Fishman, P. M. (1978). What do couples talk about when they're along? In D. Burtuff, & E. L. Epstein (Eds.), *Women's language and style* (pp. 11-22). Akron, OH: L & S Books.

Fishman, P. M. (1983). Interaction: The work women do. In B. Thorne, C. Kramarae & N. Henley (Eds.), *Language, gender, and society* (pp. 89-101). Cambridge: Newbury House.

Flekova, L., & Gurevych, I. (2013). Can we hide in the web? Large scale simultaneous age and gender author profiling in social media. *Note book for PAN at CLEF 2013*. Retrieved from http://ceur-ws.org/Vol-1179/CLEF2013wn-PAN-FlekovaEt2013.pdf

Flynn, E. A. (1983). Gender differences and student writing. Annual Meeting of the *Conference on College Composition and Communication*, Detroit, ML.

Fowler, E. F., & Ridout, T. N. (2009). Local television and newspaper coverage of political advertising. *Political Communication, 26*(2), 119-136.

Francis, W. N., & Kucera, H. (1979). *Manual of information to accompany a standard corpus of present-day edited American English, for use with digital computers*. Retrieved from http://khnt.hit.uib.no/icame/manuals/brown/index.htm

Freed, A. F. (1996). Language and gender research in an experimental setting. In V. L. Bergvall, J. M. Bing & A. F. Freed (Eds.) *Rethinking language and gender research: Theory and practice* (pp. 54-76). London: Longman.

Galimberti, C. (1994). Dalla comunicazione alla conversazione [From communication to conversation]. *Ricerche di Psicologia*, *18*(1), 113-152.

Gaman. (2007). *50 Most influential blogs in Malaysia*. Retrieved August 2, 2011, from www.sabahan.com/2007/02/06/50-most-influential-blogs-in-malaysia/

Gambell, T., & Hunter, D. (2000). Surveying gender differences in Canadian school literacy. *Journal of Curriculum Studies, 32*(5), 689-719.

Garside, R., Leech, G., & Sampson, G. (1987). *The Computational Analysis of English: A Corpus-based Approach*. London: Longman

Garside, R., & Smith, N. (1997). A hybrid grammatical tagger: CLAWS4. In R. Garside, G. Leech & A. McEnery (Eds.), C*orpus annotation: Linguistic information from computer text corpora* (pp. 102-121). London: Longman.

Gedera, D. S. P. (2012). The dynamics of blog peer feedback in ESL classroom. *Teaching English with Technology, 12*(4), 16-30.

Giddens, A. (1989). *Sociology*. Cambridge: Polity Press.

Glenn, D. (2003, June 6). Scholars who blog. *The Chronicle of Higher Education*, p. A14.

Gliozzo, A. (2005). *Semantic domains in computational linguistics*. (Unpublished doctoral dissertation). University of Trento, Trentino, Italy.

Gliozzo, A. (2006). The GOD Model. *The Eleventh Conference of the European Chapter of the Association for Computational Linguistics*, 147-150. Retrieved from http://acl-arc.comp.nus.edu.sg/archives/acl-arc-090501d3/data/pdf/anthology-PDF/E/E06/E06-2016.pdf

Godwin-Jones, B. (2003). Emerging technologies: blogs and wikis: environments for on-line collaboration. *Language Learning & Technology, 7*(2), 12-16.

Goldberg, G. L., & Roswell, B. S. (2002). *Reading, writing, and gender: Instructional strategies and classroom activities that work for girls and boys*. New York, NY: Eye on Education.

Graddol, D., & Swann, J. (1989). *Gender voices*. Oxford: Blackwell Publishers.

Guadagno, R. E., Okdie, B. M., & Eno, C. A. (2008). Who blogs? Personality predictors of blogging. *Computers in Human Behavior, 24,* 1993-2004.

Gunn, C., & McSporran, M. (2003). Dominant or different? Gender issues in computer supported learning. *Journal of Asynchronous Learning Networks, 7* (1), 14-30.

Halla, J. (2009). *Analysing user created content in blogs: A case study of fashion bloggers' PR-Trip to Stockholm.* (Unpublished master's thesis). University of Tampere, Tampere, Finland.

Halmetoja, T. (2013). *Gender-related variation in CMC language: A study of three linguistic features on twitter.* (Unpublished master's thesis). University of Gothenburg, Gothenburg, Sweden.

Hambrick, M. E., Simmons, J. M., Greenhalgh, G. P., & Greenwell, T. C. (2010). Understanding professional athletes' use of twitter: A content analysis of athletes' tweets. *International Journal of Sport Communication, 3,* 454-471.

Hamdan, S. (2011). Identifying the linguistic genderlects of the style of writing of Arab male and female novelists. *Journal of Education Culture and Society, 2,* 55-62.

Harasim, L. (1990). Introduction to Online Education. In L. Harasim (Ed.), *Online education: Perspectives on a new environment* (pp. xvii-xxiii). New York, NY: Praeger.

Hardie, A., Koller, V., Rayson, P., & Semino, E. (2007, July). *Exploiting a semantic annotation tool for metaphor analysis*. Paper presented at Corpus Linguistics Conference (CL2007), Birmingham, UK.

Hartman, M. (1976). A descriptive study of the language of men and women born in Maine around 1900 as it reflects the Lakoff hypotheses in language and woman's place. In B. L. Dubois & I. Crouch (Eds.), *The sociology of the languages of American women* (pp. 81-90). San Antonio, TX: Trinity University Press.

Hasan, J. M., & Khammat, A. H. (2011). Gender differences in the use of sentence type and sentence modifiers: An analysis of emails by Iraqi learners of English. *Journal of College of Arts, 57,* 1-17.

Herring, S. C. (1992a, January). *Gender and participation in computer-mediated linguistic discourse*. Paper presented at the Annual Meeting of the Linguistic Society of America, Philadelphia, USA.

Herring, S. C. (1992b). Men's language: A study of the discourse of the LINGUIST list. In A. Crocheti ère, J-C. Boulanger & C. Ouellon (Eds.), *Endangered languages:*

Proceedings of the 15th International Congress of linguists (pp. 347-350). Quebec: Les Presses de l'Université Laval.

Herring, S. C. (1993). Gender and democracy in computer-mediated communication. *Electronic Journal of Communication, 3*(2), 1-17.

Herring, S. C. (1994, June). *Gender differences in computer-mediated communication: bring familiar baggage to the new frontier*. Paper presented at the American Library Association Annual Convention. Miami, USA.

Herring, S. C. (2000). Gender differences in CMC: Findings and implications. *Computer Professionals for Social Responsibility Journal, 18* (1). Retrieved January 5, 2005 from http://cpsr.org/prevsite/publications/newsletters/issues/2000/Winter2000/herring.html/

Herring, S. C. (2001). Computer-mediated discourse. In D. Schiffrin, D. Tannen & H. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 612-637). Oxford: Blackwell.

Herring, S. C., Kouper, I., Scheidt, L. A., & Wright, E. (2004). Women and children last: The discursive construction of weblogs. In L. Gurak, S. Antonijevic, L. Johnson, C. Ratliff & J. Reyman (Eds.), *Into the blogosphere: Rhetoric, community, and culture of weblogs.* University of Minnesota. Retrieved from http://blog.lib.umn.edu/blogosphere/women_and_children.html

Herring, S. C., Scheidt, L. A., Bonus, S., & Wright, E. (2004). Bridging the gap: A genre analysis of weblogs. Proceedings of the *37th Hawaii International Conference on System Sciences* (pp. 101-111). Los Alamitos: IEEE Computer Society Press.

Herring, S. C., & Paolillo, J. C. (2006). Gender and genre variation in weblogs. *Journal of Sociolinguistics, 10*(4), 439-459.

Herring, S. C., & Zelenkauskaite, A. (2009). Symbolic capital in a virtual heterosexual market: abbreviation and insertion in Italian iTV SMS. *Written Communication, 26*(1), 5-31.

Herzog, H. A. (2007). Gender differences in human-animal interactions: A review. *Anthrozoös, 20*(1), 7-21.

Heuer, C. A., McClure, K. J., & Puhl, R. M. (2011). Obesity stigma in online news: A visual content analysis. *Journal of Health Communication, 0*, 1-12.

Hock, D. (1999). *Birth of the chaodic age*. San Francisco: Barrett-koehler.

Holmes, J. (1990). Hedges and boosters in women's and men's speech. *Language & Communication, 10*(3), 185-205.

Holmes, J., & Stubbe, M. (2003). *Power and politeness in the workplace: A sociolinguistic analysis of talk at work*. London: Longman.

Hopkins, J. (2010). *Chronology of English blogs in Malaysia*. Retrieved from http://julianhopkins.net/uploads/jh_100107_BlogHistoryChronology.pdf

Huff, C. A. (1996). Textual boundaries: Space in Nineteenth-century women's manuscript dairies. In S. L. Bunkers & C. A. Huff (Eds.), *Inscribing the daily: Critical essays on women's diaries* (pp. 123-138). Amherst: University of Massachusetts Press.

Huffaker, D. (2004). *Gender similarities and differences in online identity and language use among teenage bloggers*. (Unpublished master's thesis). Georgetown University, Washington, D. C., USA.

Huffaker, D., & Calvert, S. L. (2005). Gender, identity and language use in teenage Blogs. *Journal of Computer-Mediated Communication, 10*(2). Retrieved from http://jcmc.indiana.edu/vol10/issue2/huffaker.html

Izquierdo, L., & Reyes, E. (2009). Effectiveness of blogging to practice reading at a freshman EFL program. *Reading matrix: An international online journal, 9*(2), 100-117.

James, D., & Drakich, J. (1993). Understanding gender differences in amount of talk. In D. Tannen (Ed.), *Gender and conversational interaction* (pp. 281-312). New York, NY: Oxford University Press.

James, K. E., John F., & Hensel, P. J. (1995). Comparative magazine advertisements revisited: a content analysis. In B. T. Engelland & D. T. Smart (Eds.), in *Proceedings of the Annual Meeting of the Southern Marketing Association* (pp. 109-113). Southern Marketing Association, Evansville, Indiana.

Jawhar, S. S. (2012). *Conceptualising CLIL in a Saudi context: a corpus linguistic and conversation analytic perspective*. (Unpublished doctoral dissertation). Newcastle University, Newcastle, UK.

Jayaraman, K., Munira, H., Dababrata, C., & Iranmanesh, M. (2013). The preference and consumption of chicken lovers with race as a moderator – an empirical study in Malaysia. *International Food Research Journal, 20*(1), 165-174.

Johnstone, B. (1993). Community and contest: Midwestern men and women constructing their worlds in conversational storytelling. In D. Tannen (Ed.), *Gender and conversational interaction* (pp. 62-80). Oxford: Oxford UP.

Jespersen, O. (1922). The woman. In D. Cameron, (Ed.), *The feminist critique of language* (pp. 225-241). London: Routledge.

Jones, S., & Myhill, D. (2007). Discourses of difference? Examining gender differences in linguistic characteristics of writing. *Canadian Journal of Education, 30*(2), 456-482.

Kann, C. L. (2001). *The effects of gender on Internet-assisted English writing instruction for senior high school students.* (Unpublished master's thesis). National Kaohsiung Normal University, Taiwan.

Kassam, M. H. (1996). Some aspects of women's voices from Northern Nigeria. *African Languages and Cultures, 9*(2), 111-125.

Kaur, S. (2009). *A corpus-driven contrastive study of girls' and boys' use of vocabulary in their writing in Malaysia and the United Kingdom*. (Unpublished doctoral dissertation). Lancaster University, Lancaster, UK.

Kaur, S., Ganapathy, M., & Sidhu, G. K. (2012). Designing learning elements using the multiliteracies approach in an ESL writing classroom. *3L: The Southeast Asian Journal of English Language Studies*, *18*(3), 119-134.

Kavanagh, B. (2010). A cross-cultural analysis of Japanese and English non-verbal online communication: The use of emoticons in weblogs. *Intercultural Communication Studies XIX, 3*, 65-80.

Kent, M. (2008). Critical analysis of blogging in public relations. *Public Relations Review, 34*(1), 32-40.

Key, M. R. (1975). *Male/Female language*. Metuchen, NJ: The Scarecrow Press.

Kim, S. (2009). Content analysis of cancer blog posts. *Journal Medical Library Association, 97*(4), 260-266.

Kinzie, S. (2005, March 11). Blogging clicks with colleges: Interactive web pages changing class participation. *The Washington Post*, p. B01.

Kirchhoff, L., Bruns, A., & Nicolai, T. (2007, October). *Investigating the impact of the blogosphere: Using PageRank to determine the distribution of attention.* Paper

presented at the Association of Internet Researchers Conference, Vancouver, Canada.

Knowles, G., & Zuraidah Mohd Don. (2004, October). *Introducing MACLE: The Malaysian corpus of learner English*. Paper presented at the First National Symposium of Corpus Linguistics and Foreign Language Education Conference, Guangzhou, China.

Koller, V., Hardie, A., Rayson, P., & Semino, E. (2008). Using a semantic annotation tool for the analysis of metaphor in discourse. *Metaphorik.de, 15*, 141-60.

Koppel, M., Argamon, S., & Shimoni, A. R. (2002). Automatically categorizing written texts by author gender. *Literary and Linguistic Computing 17*(4), 401-412.

Kozma, R. (1991). Learning with media. *Review of Educational Research, 61*(2), 179-211.

Krennmayr, T. (2011). *Metaphor in newspapers*. Utrecht: LOT.

Lakoff, T. R. (1975). Language and woman's place. In M. Bucholtz, (Ed.), *Language and woman's place: Text and commentaries* (pp. 35-102). Oxford: Oxford University Press.

Lakoff, T. R. (2004). Language and woman's place. In M. Bucholtz, (Ed.), *Revised and expanded edition* (pp. 78-81). Oxford: Oxford University Press.

Lee, A. (1996). *Gender, literacy, curriculum: Re-writing school geography*. London: Taylor & Francis.

Leech, G. (1992). Corpora and theories of linguistic performance. In J. Svartvik (Ed.), *Directions in corpus linguistics: Proceedings of Nobel Symposium 82* (pp. 105-122). Berlin: Mouton de Gruyter.

Leech, G. (1997). Introducing corpus annotation. In R. Garside, G. Leech & A. McEnery (Eds.), *Corpus annotation: Linguistic information from computer text corpora* (pp. 1-18). London: Longman.

Leech, G., & Fallon, R. (1992). Computer corpora-What do they tell us about culture? *I-CAME Journal, 16,* 29-50.

Lei, X. L. (2006). Sexism in language. *Journal of Language and Linguistics, 5*(1), 87-94.

Lenhart, A., & Fox, S. (2006). Bloggers: A portrait of the internet's new storytellers. *Pew Internet & American Life Project.* Retrieved from

http://www.pewinternet.org/pdfs/PIP%20Bloggers%20Report%20July%2019%202006.pdf

Levine, T., & Geldman-Caspar, Z. (1996). Informal science writing produced by boys and girls: Writing preference and quality. *British Educational Research Journal, 22*(4), 421-439.

Li, D. (2005). *Why do you blog: A uses-and-gratifications inquiry into bloggers' motivations*. (Unpublished master's thesis). Marquette University, Milwaukee, USA.

Li, D. J., An, S. H., & Yang, H. (2007). A content analysis of the newspaper advertisements westernization in China. *Frontiers of Business Research in China, 1*(4), 558-573.

Li, Q. (2005, April). *Cyberbullying in schools: Nature and extent of adolescents' experience.* Paper presented at the Annual American Educational Research Association Conference, Montreal, Canada.

Llach, M. P. A., & Gallogo, M. T. (2012). Vocabulary knowledge development and gender differences in a second language. *Estudios de Lingüística Inglesa Aplicada (ELIA), 12*, 45-75.

Long, P. (2002). *Blogs: A disruptive technology coming of age? Syllabus.* Retrieved January 15, 2005, from http://www.campustechnology.com/article.asp?id=6774

Lord, G., & Lomicka, L. L. (2004). Developing collaborative cyber communities to prepare tomorrow's teachers. *Foreign Language Annals, 37*(3), 401-417.

Lövheim, M. (2011). Personal and popular: The case of young Swedish female top-bloggers. *Nordicom Review, 32*(1), 3-16.

Lyons, J. (1977). *Semantics*. Cambridge: Cambridge University Press.

Maclaran, P., & Miriam, C. (2002). Analysing qualitative data: Computer software and the market research practitioner. *Qualitative Market research: An International Journal 5*(1), 28-39.

Magnini, B., Strapparava, C., Pezzulo, G., & Gliozzo, A. (2002). The role of domain information in word sense disambiguation. *Natural Language Engineering, 8*(4), 359-373.

Mah, B. Y., & Er, A. N. (2009). Writing web logs in the ESL classroom: A study of student perceptions and the Technology Acceptance Model. *Asian Journal of University Education (ACRULeT), 5*(1), 47-70.

Mah, B. Y., & Liaw, S. C. (2011). Are you ready to blog? In F. Mohamad (Ed.), *Technology in ESL classrooms* (pp. 33-54). Shah Alam, Selangor: UTM Publication Centre.

Marshall, I. (1983). Choice of grammatical word-class without global syntactic analysis: Tagging words in the LOB corpus. *Computers and the Humanities, 17,* 139-150.

McArthur, T. (1981). *Longman lexicon of contemporary English*. London: Longman.

Mccluskey, M. (2005). *A content analysis of 2004 presidential election headlines of The Los Angeles Times and The Washington Times*. (Unpublished master's thesis). University of Central Florida, Orlando, USA.

McConnell-Ginet, S. (1978). Intonation in a man's world. *Signs: Journal of Women in Culture and Society, 3,* 541-559.

McEnery, T., & Wilson, A. (2001). *Corpus linguistic* (2nd ed.). Edinburgh, Scotland: Edinburgh University Press.

McGinnis, T., Goodstein-Stolzenberg, A., & Saliani, E. C. (2007). "Indnpride": Online spaces of transnational youth as sites of creative and sophisticated literacy and identity work. *Linguistics and Education, 18,* 283-304.

McMenamin, I., Flynn, R., O'Malley, E., & Rafter, K. (2013). Commercialism and election framing: A content analysis of twelve newspapers in the 2011 Irish general election. *The International Journal of Press/Politics, 18*(1), 3-20.

McNamara, K. R. (2007). *Blogging breast cancer: language and subjectivity in women's online illness narratives*. (Unpublished doctoral dissertation). Georgetown University, Washington, D. C., USA.

Moon, D., & Lim, D. (2013). Using weblogs for foreign language learning: Promises and challenges. *Proceedings of the Seventh International Conference on Information Security and Assurance, 21,* 208-212. Retrieved from http://onlinepresent.org/proceedings/vol21_2013/53.pdf

Morgan, H. C., Banks, J., & Boals, A. (2011). Social anxiety and word use: How environments can influence words. *The Journal of Young Investigators, 21*(5), 71-83.

Morris, P. K. & Nichols, K. (2013). Conceptualizing beauty: A content analysis of U.S. and French women's fashion magazine advertisements. *Online Journal of Communication and Media Technologies, 3*(1), 49-74.

Mulac, A., Studley. L., & Blau, S. (1990). The gender-linked language effect in primary and secondary students' impromptu essays. *Sex Roles, 23*(9/10), 439-470.

Mulac, L., & Lundell, T. L. (1994). Effects of gender-linked language differences in adults' written discourse: Multivariate tests of language effects. *Language & Communication, 14*(3), 299-399.

Murty, D. V. R. (2001). *Developmental journalism*. New Delhi, India: Dominant Publishers.

Mustapha, R., & Wong, L. S. M. (2012). Language use among Malaysian bloggers. (2012). *Atikan, 2*(1). Retrieved from http://atikan-jurnal.com/wp-content/uploads/2012/06/7.ramlee.mei_.upsi_.jun_.12.pdf

Nahon, K., & Hemsley, J. (2011, January). *Democracy.com: A tale of political blogs and content.* Paper presented at the 44th Hawaii International Conference on System Sciences. Manoa, USA.

Nemati, A., & Bayer, J. M. (2007). Gender differences in the use of linguistic form in the speech of men and women: A comparative study of Persian and English. *Language in India, 7,* 1-16.

Nepomuceno, M. M. (2011). Writing online: Using blogs as an alternative writing activity in tertiary ESL Classes. *TESOL Journal, 5,* 92-105.

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. D*iscourse Processes, 45,* 211-236.

Ning, H. L., Dai, X., & Zhang, F. C. (2010). On gender difference in English Language and its causes. *Asian Social Science, 6*(2), 126-130.

Nizam, D. N. M., Yee, C. P., Yussof, S., Belongkikit, J., & Ibrahim, A. A. A. (2011). Gender disparity on blogging awareness among Malaysian tertiary students. *2011 International Conference on Research and Innovation in Information Systems* (pp. 23-24). Kuala Lumpur: ICRIIS' 11.

Nor Ida Ramli (2005). *Corpus-based approaches to investigating similar words in Malay*. (Unpublished doctoral dissertation). Lancaster University, Lancaster, UK.

Nowson, S. (2006). *The language of weblogs: A study of genre and individual differences.* (Unpublished doctoral dissertation). University of Edinburgh, Edinburgh, UK.

Nowson, S., & Oberlander J. (2006, March). *The identity of bloggers: Openness and gender in personal weblogs*. Paper presented at AAAI Spring Symposium, Computational Approaches to Analyzing Weblogs, Stanford, CA, USA.

Nowson, S., & Oberlander, J. (2007, March). *Identifying more bloggers: Towards large scale personality classification of personal weblogs*. Paper presented at the International Conference on Weblogs and Social Media. Boulder, CO, USA.

Ooi, V. B. Y., Tan, P. K. W., & Chiang, A. K. L. (2007). Analyzing personal weblogs in Singapore English: The WMatrix approach. *Journal of the Research Unit for Variation, Contacts, and Change in English, 2*. Retrieved from http://www.helsinki.fi/varieng/series/volumes/02/ooi_et_al/

Palander-Collin, M. (1999). Male and female styles in 17th century correspondence. *Language Variation and Change, 11,* 123-141.

Pan, B., MacLaurin, T., & Crotts, J. C. (2006). Travel blogs and the implications for destination marketing. *Journal of Travel Research, 46*(1), 35-45.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*, 547-577.

Peterson, S. (2002). Gender meanings in grade eight students' talk about classroom writing. *Gender and Education, 14*(4), 351-366.

Pollock, N., Okdie, B. M., & Guadagno, R. E. (2011). Gender difference in religious as a means of self-enhancement: The impact of perceived audience size and identifiablity. *The Yale Review of Undergraduate Research in Psychology, 1*, 45-49.

Poole, M. E. (1979). Social class, sex, and linguistic coding. *Language and Speech, 22*, 49-67.

Porter, L., Sweetser, K., & Chung, D. (2009). The blogosphere and public relations: Investigating practitioners' roles and blog use. *Journal of Communication Management, 13*(3), 250-267.

Prasad, B. D. (2008). Content analysis: A method in social science research. In D. K. Lal Das & V. Bhaskaran (Eds.), *Research methods for social work* (pp. 173-193). New Delhi: Rawat.

Preisler, B. (1986). *Linguistic sex roles in conversation*. Berlin: Mouton de Gruyter.

Puhl, R. M., Moss-Racusin, C. A., & Schwartz, M. B. (2007). Internalization of weight bias: Implications for binge eating and emotional well-being. *Obesity, 15*(1), 19-23.

Quicke, K. (2012). "I'm a big manly man!" A study of embedded gender stereotypes in video games. *Griffith Working Papers in Pragmatics and Intercultural Communication, 5*(1), 1-13.

Rangel, F., & Rosso, P. (2013). Use of language and author profiling: Identification of gender and age. In the Proceeding of *10th International Workshop on Natural Language Processing and Cognitive Sciences* (pp. 13-17), Marseille: NLPCS 2013 CIRM.

Rayson, P., Berridge, D., & Francis, B. (2004, March). *Extending the Cochran rule for the comparison of word frequencies between corpora*. Paper presented at the Seventh International Conference on Statistical analysis of textual data (JADT 2004) (pp. 926-936), Louvain-la-Neuve, Belgium.

Rayson, P., Leech, G., & Hodges, M. (1997). Social differentiation in the use of English vocabulary: Some analyses of the conversational component of the British National Corpus. *International Journal of Corpus Linguistics, 2*, 133-152.

Rayson, P. (2003). W*Matrix: A statistical method and software tool for linguistic analysis through corpus comparison.* (Unpublished doctoral dissertation). Lancaster University, Lancaster, UK.

Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics, 13*(4), 519-549.

Rayson, P. (2013). *WMatrix corpus analysis and comparison tool*. Computing Department, Lancaster University. Retrieved from http://ucrel.lancs.ac.uk/wmatrix/

Reese, S., Rutigliano, L., Hyun, K., & Jeong, J. (2007). Mapping the blogosphere: Citizen-based media in the global news arena. *Journalism: Theory, Practice, Criticism, 8*(3), 235-262.

Richardson, W. (2005). *Blogs, wikis, podcasts, and other powerful web tools for classrooms*. Thousand Oaks, CA: Corwin Press.

Risager, K. (2005). Languaculture as a key concept in language and culture teaching. In A. F. Bent Preisler, S. Kjaerbeck & K. Risager (Eds.), *The consequences of mobility* (pp. 185-196). Roskilde: Roskilde University.

Ritzer, G. (2004). *Encyclopaedia of social theory* (vols. 1-2). Thousand Oaks, CA: Sage.

Roberts, S. (2003). Campus communications and the wisdom of blogging. *Campus Technology*. Retrieved January 4, 2007, from http://www.campustechnology/article.aspx?aid=39436

Robinson, T., & Callister, M. (2008). Body image of older adults in Magazine advertisements: A content analysis of their body shape and portrayal. *Journal of Magazine and New Media Research, 10*(1), 1-16.

Rodino, M. (1997). Breaking out of binaries: Reconceptualizing gender and its relationship to language in computer-mediated communication. *Journal of Computer-Mediated Communication, 3*(3), 0. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.1997.tb00074.x/full

Rodzvilla, J. (2002). *We've got blog: How weblogs are changing our culture*. Cambridge, MA: Perseus Publishing.

Saito, S. (2007). Television and the cultivation of gender-role attitudes in Japan: Does television contribute to the maintenance of the Status Quo? *Journal of Communication, 57,* 511-531.

Santosh, K., Bansal, R., Shekhar, M., & Varma, V. (2013). Author profiling: Predicting age and gender from blogs. *Notebook for PAN at CLEF.* Retrieved from http://www.uni-weimar.de/medien/webis/research/events/pan-13/pan13-papers-final/pan13-author-profiling/santosh13-notebook.pdf

Schafferer, C. (2004). Content analysis of newspaper and television advertisements: A case study of Taiwan's 2004 presidential election. *Modern East Asia, 3*(4), 1-19.

Scheidt, L. A. (2006). Adolescent diary weblogs and the unseen audience. In D. Buckingham & R. Willett (Eds.), *Digital generations: Children, young people and new media* (pp. 193-210). London: Lawrence Erlbaum.

Schler, J., Koppel, M., Argamon, S., & Pennebaker, J. (2006). Effects of age and gender on blogging. *Proceedings of 2006 AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs,* 199-205.

Schmid, H. (2003). Do women and men really live in different cultures? Evidence from the BNC. In A. Wilson, P. Rayson & T. McEnery (Eds.), *Corpus linguistics by the Lune: a festschrift for Geoffrey Leech* (pp. 185-221). Frankfurt: Perter Lang.

Schultz, P. (2013). Gender variation in Writing: Analyzing online dating ads. *Coyote Papers 21*. Retrieved from http://arizona.openrepository.com/arizona/bitstream/10150/271015/1/schultz.pdf

Scott, M. (1999). *WordSmith: software tools for Windows* (Vols. 1-3). Oxford: University Press.

Segev, S., Villar, M. E., & Fiske, R. M. (2012). Understanding opinion leadership and motivations to blog: Implications for public relations practice. *Public Relations Journal*, *6*(5). Retrieved from http://www.prsa.org/Intelligence/PRJournal/Documents/2012Segev.pdf

Seitzinger, J. (2006). Be constructive: blogs, podcasts, and wikis as constructivist learning tools. *Learning Solutions e-Magazine*, *7*. Retrieved from http://www.careersmarts.com/21/blogs%20wikis%20podcasts%20learning%20tools.pdf

Selfe, C. L., & Paul, R. M. (1991). Testing claims for on-line conferences. *Written Communication, 8*(2), 163-192.

Selingo, J. (2004, August 19). In the classroom, weblogs are the new bulletin boards. *New York Times*, p. G7.

Sharp, G. (2012). That is *so* feminine! An investigation of intensifiers as characteristics of female speech through the use of *so* and *really* in modern television programming. *Griffith Working Papers in Pragmatics and Intercultural Communication, 5*(1), 14-20.

Simpson, A. P. (2003). Possible articulatory reasons for sex-specific differences in vowel duration. *Proceeding of 6th International Seminar on Speech Production*, 261-266.

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. London: Routledge.

Singh, G., & Shahid, S. M. (2006). Blogs as tools for online information sharing. *Library Philosophy and Practice*, *9*(1), Paper 96. Retrieved from http://www.webpages.uidaho.edu/~mbolin/singh-shahid.pdf

Skoglund, J. (2009). Gender difference in role-play: Male and female character language in world of Warcraft. *English C/Special Project.* Retrieved from http://www.diva-portal.org/smash/get/diva2:158742/FULLTEXT01.pdf

Smeltzer, S. (2008) Blogging in Malaysia: Hope for a new democratic technology? *Journal of International Communication, 14*(1), 28-45.

Smith, B. G. (2011). Becoming "quirky:" Towards an understanding of practitioner and blogger relations in public relations. *Public Relations Journal, 5*(4), 1-17.

Smith, J., & Balka, E. (1988). Chatting on a feminist network. In C. Kramarae (Ed.), *Technology and women's voices* (pp. 82-97). New York: Routledge and Kegan Paul.

Smith, M. N. (2011). *Amplifying the female voice in online political discourse*. (Unpublished master's thesis). American University, Washington, D. C., USA.

Spender, D. (1980). *Man made language*. Oxford: Routledge & Kegan Paul.

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.

Suad, A. (1999). *Multi word units in a corpus-based study of memoranda of understanding: Modal multi word units.* (Unpublished doctoral dissertation). Lancaster University, Lancaster, UK.

Sun, X. N. (2008). *Why gender matters in CMC: Gender differences in remote trust and performance with initial social activities*. (Unpublished doctoral dissertation). Drexel University, Philadelphia, USA.

Sunderland, J. (2006). *Gender and language: An advanced resource book*. London: Routledge.

Sutton, L. (1994). Using Usenet: Gender, power, and silence in electronic discourse. In S. Gahl, A. Dolbey & C. Johnsons (Eds.), *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society* (pp. 506-520). Berkeley, CA: Berkeley Linguistics Society.

Tang, H. W. (2005, April). *Let a hundred flowers bloom: A Malaysian case study on blogging towards a democratic culture.* Paper presented at the 20th British and Irish Law Education and Technology Association Conference, Belfast, UK.

Tannen, D. (1990). *You just don't understand*. New York, NY: Ballantine Books.

Thayalan, X., & Noor, F. N. M. (2010). Gender-based strategies in Malaysian student blog. *Proceedings of 2010 International Conference on Science and Social Research (CSSR 2010),* 896-901.

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

Tossell, C. C., Kortum, P., Shepard, C. Barg-Walkow, L. H., Rahmati, A., & Zhong, L. (2012). A longitudinal study of emoticon use in text messaging from smartphones. *Computers in Human Behavior, 28*, 659-663.

Trammell, K. D. (2004). Celebrity blogs: *Investigation in the persuasive nature of two-way communication regarding politics*. (Unpublished doctoral dissertation). University of Florida, Florida, USA.

Tseng, M. C. (2008). The use of blogs in English class for medicine-related majors. *Chang Gung Journal of Humanities and Social Sciences, 1*(1), 167-187.

Turkle, S. (2004, January 30). How computers change the way we think. *Chronicle of Higher Education*, pp. B26-B28.

Ubaidullah, N. H., Mahadi, N., & Ching, L. H. (2013). Exploring the educational benefits of blogs to help non-Malay pupils in Malay language learning. *World of Computer Science and Information Technology Journal, 3*(1), 20-25.

Van Gelder, L. (1990). The strange case of the electronic lover. In G. Gumpert & S. L. Fish (Eds.), *Talking to strangers: Mediated therapeutic communication* (pp. 128-142). Norwood, NJ: Ablex.

Walker, L. (2004, November 4). Bloggers gain attention in 2004 election. *The Washington Post*, p. E01.

Walsh, C. (2001). *Gender and discourse: Language and power in politics, the church and organizations*. Harlow, England: Longman.

Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research, 19*, 52-90.

Wang, S., & Vasquez, C. (2012). Web 2.0 and second language learning: What does the research tell us? *The Journal of the Computer Assisted Language Instruction Consortium, 29*(3), 412-430.

Wang, Y. S., Lin, H. H., & Liao, Y. W. (2010). Investigating the individual difference antecedents of perceived enjoyment in the acceptance of blogging. *World Academy of Science, Engineering and Technology, 43,* 1014-1023.

Wenger, A. (2008). Analysis of travel bloggers' characteristics and their communication about Austria as a tourism destination. *Journal of Vacation Marketing, 14*(2), 160-176.

Widjaja, C. S. (1997). A study of date refusal: Taiwanese females vs. American females. *University of Hawaii Working Papers in ESL, 15*(2), 1-43.

Winer, D. (2003). What makes a weblog a weblog? *Weblogs at Harvard Law*. Retrieved from: http://blogs.law.harvard.edu/whatMakesAWeblogAWeblog

Wodak, R. (2009). Critical discourse analysis: History, agenda, theory, and methodology. In R. Wodak & M. Meyer (Eds.), *Methods for critical discourse analysis*. (pp. 1-33). London: Sage.

Xia, W. J. (2012). Study on gender differences in English. *Lecture Notes in Information Technology, 16-17*, 218-221.

Xifra, J., & Huertas, A. (2008). Blogging PR: An exploratory analysis of public relations weblogs. *Public Relations Review, 34*(3), 269-275.

Yale, R. (2007). *Gender differences in instant messaging*: *Quantitative analysis of linguistic characteristics*. Saarbrücken: VDM Verlag Dr. Müller.

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality, 44*, 363-373.

Yip, S. M., (2006). *The gender differences in linguistic features of dyadic synchronous computer-mediated communication*. (Unpublished master's thesis). The Chinese University of Hong Kong, Hong Kong, China.

Yu, B. (2013). Language and gender in congressional speech. *Literary and Linguistic Computing.* Retrieved from http://ischool.syr.edu/media/documents/2013/2/beiyujan2013.pdf

Yunus, M. M., Kiing, J. L., & Salehi, H. (2013). Using blogs to promote writing skill in ESL classroom. *Proceedings of the 4th International Conference on Education and Educational Technologies* (EET '13), 109-113.

Zare, P. (2012). Language learning strategies among EFL/ESL learners: A review of literature. *International Journal of Humanities and Social Science, 2*(5), 162-169.

APPENDICES

Appendix 1: Sample from Female Weblogs

Thursday, December 20, 2012
20-12-2012!

Happy 20-12-2012!

The year is coming to a close and all my travel memories are also coming to an end. I will most likely continue with my travel adventures in Monaco and Holland next year. For now, I am taking things easy, tying up loose ends at work and at home, all the while wishing that another adventure someplace awesome will come my way soon.

Window shopping in Monte Carlo

Have you made your New Year resolutions yet? I know some of you don't like making resolutions but I tend to make a few every year, which I've posted at my perfume blog. Hope to stick to them in order to ensure 2013 remains drama-free, easygoing and fruitful for me and those close to my heart.

Meanwhile, here are some awesome tips for the happy travellers out there. They're pretty handy to take note prior to your trip and I will definitely keep this travel guide in mind for my next getaway, wherever and whenever that will be!

Appendix 2: Sample from Male Weblogs

Monday, February 4, 2013

Sorrento: Barbeque Feast

Sorrento will always have a special place in my heart. It will forever be my favourite European city with the general Tuscany region coming in a close second. Us boys were in Sorrento for a few days, the second stop on our tour around Europe last summer. We stayed on a campsite located on the top of the Amalfi Coast, just a 20 minute walk away from the town centre. The camp has it's own beach down the hill. We usually go there in the scorching afternoons to cool off as the cabins tend to get a little stuffy then.

While exploring the town on day I noticed a few shops selling fresh seafood products. They were closed at the time and using the very limited amount of Italian I knew, found out from the locals that I would have to come early in the morning. My friends would know how much I struggle with mornings. But wake up I did. It was our final night in Sorrento and we agreed to split the cost with two other friendly travellers we met, Ana and Matthew.

I dragged myself down to the market that morning begrudgingly. There was this truck outside the shop unloading the seafood caught by fisherman early that morning. I started becoming really excited and made my way around the shop pointing at things. Luckily they use the same measuring system so it made it easier to tell them how much I want. A good 45 minutes later I exited the shop triumphantly with three bags of fresh seafood filled to the brim with mussels, prawns, two fishes. Got them for such good value. I think I got 2kg of mussels for about 5 euros. Everything came to about thirty-ish.

We cooked everything over the fire. I stuck to what I've learnt from years of watching the food and travel channels on tv. The common thing that everyone could agree on, from Nigella to Jamie is that Italian cooking is meant to be kept simple, allowing the core ingredients to take the limelight. The Italians take pride in the quality of their products and we aim to retain that by not over-complicating the cooking process.

Garlic, parsley, salt and olive oil. That was all we used. Well, except for the occasional cheeky splash of white wine. Dinner was amazing. It's hard to go wrong when you have such good ingredients and basic cooking instructions. We ate and drank our fill, way into the night. I still have dreams about those prawns. Their heads were bursting with flavourful juices and you can suck on them all night. This is how I imagine life to be after retirement. Bliss.

Appendix 3: List of Mistags in the Corpora (Chapter 4)

Malaysian female key semantic domains:

I. Mistags in the Malaysian female weblogs (Reference Table 4.2)

No.	Tag	Semantic Domain	Mistag	Total recalculated
1	F1	Food	-	636
2	Z8	Pronouns	-	5567
3	B3	Medicines and medical treatment	-	165
4	S4	Kin	-	282
5	S2.1	People: Female	*miss* (1)- verb	68-1=67
6	B1	Anatomy and physiology	-	363
7	B2-	Disease	-	137
8	H2	Parts of buildings	-	82
9	T3--	Time: New and young	-	38
10	C1	Arts and crafts	-	153
11	L2	Living creatures: animals, birds, etc.	*bunny* (2), *tiger* (3)- proper noun; *fish* (1)- verb	143-6=137
12	O4.4	Shape	*straight* (3)- adverb	69-3=66

Malaysian female key semantic domains:

II. Mistags in the Malaysian male weblogs (Reference Table 4.2)

No.	Tag	Semantic Domain	Mistag	Total recalculated
1	F1	Food	-	343
2	Z8	Pronouns	-	4529
3	B3	Medicines and medical treatment	*sen* (4)- monetary unit	69- 4=65
4	S4	Kin	-	145
5	S2.1	People: Female	*miss* (1)- verb	15- 1=14
6	B1	Anatomy and physiology	*physical* (2)- real, *shit* (6)- taboo word	232- 8=224
7	B2-	Disease	-	62
8	H2	Parts of buildings	-	28
9	T3--	Time: New and young	-	9
10	C1	Arts and crafts	*forge* (4)- proper noun	87- 4=83
11	L2	Living creatures: animals, birds, etc.	*bear* (1)- proper noun, *groom* (5)- verb	78- 6=72
12	O4.4	Shape	*straight* (2)- adverb	27-2=25

Malaysian male key semantic domains:

I. Mistags in the Malaysian male weblogs (Reference Table 4.4)

No.	Tag	Semantic Domain	Mistag	Total recalculated
1	Z2	Geographical names	-	481
2	A11.1+	Important	-	82
3	W3	Geographical terms	*sea* (2)- proper noun, *channels* (2)- TV, *atmosphere* (1)- general impression	96- 5=91
4	N3.5	Measurement: Weight	-	46
5	G3	Warfare, defence and the army; weapons	-	74
6	M7	Places	-	177
7	Z5	Grammatical bin	-	10648
8	I1.1+	Money: Affluence	-	21
9	K2	Music and related activities	-	68
10	I4	Industry	-	25
11	N5+	Quantities: many/much	-	205
12	I1.1	Money and pay	*Seoul* (10)- proper noun, *capital* (1)- location of government, *save* (1)- help	85- 12=73
13	S1.1.3+	Participating	-	41
14	Y2	Information technology and computing	*program* (8)- plan	140- 8=132
15	A3+	Existing	-	1384
16	S7.3	Competition	-	9
17	A6.1	Comparing: Similay/different	-	18
18	S9	Religion and the supernatural	*nt* (29)- not	183- 29=154
19	X2.6+	Expected	*due* (3)- due to	87- 3=84
20	A6.2+	Comparing: Usual	-	94

Malaysian male key semantic domains:

II. Mistags in the Malaysian female weblogs (Reference Table 4.4)

No.	Tag	Semantic Domain	Mistag	Total recalculated
1	Z2	Geographical names	*la* (18), *lah* (9)- modal particle	328- 27=301
2	A11.1+	Important	-	26
3	W3	Geographical terms	*atmosphere* (1)- general impression	33- 1=32
4	N3.5	Measurement: Weight	-	9
5	G3	Warfare, defence and the army; weapons	-	28
6	M7	Places	-	105
7	Z5	Grammatical bin	-	10592
8	I1.1+	Money: Affluence	-	2
9	K2	Music and related activities	-	29
10	I4	Industry	-	4
11	N5+	Quantities: many/much	*couple* (2)- two married people	145- 2=143
12	I1.1	Money and pay	*Kitty* (8), Putrajaya (1), *Isa* (1)- proper noun, *credits* (1)- praise	48- 11=37
13	S1.1.3+	Participating	-	15
14	Y2	Information technology and computing	*program* (5)- plan	90- 5=85
15	A3+	Existing	-	1275
16	S7.3	Competition	-	0
17	A6.1	Comparing: Similay/different	-	3
18	S9	Religion and the supernatural	*nt* (40)- not, *lama* (3)- Malay language	150- 43=107
19	X2.6+	Expected	*due* (2)- due to, *awaits* (1)- wait	52- 3=49
20	A6.2+	Comparing: Usual	-	57

Appendix 4: Wmatrix Key Semantic Domains in the Corpora (Chapter 4)

Key semantic domains in the Malaysian female weblogs: (Section 4.3.1)

Key semantic domain	Female weblogs	Male weblogs
1. People and relationships		
Pronouns (Z8)	1. I (1392, 3.35%) 2. my (586, 1.41%) 3. it (548, 1.32%) 4. you (310, 0.75%) 5. that (288, 0.69%) 6. we (283, 0.68%) 7. me (268, 0.64%) 8. he (166, 0.40%) 9. her (147, 0.35%) 10. they (143, 0.34%)	1. I (1133, 2.89%) 2. it (466, 1.19%) 3. my (418, 1.07%) 4. you (349, 0.89%) 5. that (299, 0.76%) 6. we (236, 0.60%) 7. me (222, 0.57%) 8. your (123, 0.31%) 9. our (94, 0.24%) 10. he (94, 0.24%)
Kin (S4)	1. family (31, 0.07%) 2. brother (21, 0.05%) 3. father (21, 0.05%) 4. mom (17, 0.04%) 5. sister (16, 0.04%) 6. husband (14, 0.03%) 7. hubby (12, 0.03%) 8. son (10, 0.02%) 9. parents (9, 0.02%) 10. grandma (8, 0.02%)	1. family (16, 0.04%) 2. wife (12, 0.03%) 3. marriage (8, 0.02%) 4. parents (8, 0.02%) 5. son (7, 0.02%) 6. daughter (7, 0.02%) 7. father (6, 0.02%) 8. wedding (6, 0.02%) 9. aunt (5, 0.01%) 10. mum (5, 0.01%)
People: Female (S2.1)	1. girl (17, 0.04%) 2. girls (14, 0.03%) 3. lady (13, 0.03%) 4. woman (6, 0.01%) 5. women (4, 0.01%) 6. ladies (2) 7. Ms (2) 8. all girls (1) 9. female (1) 10. Miss (1)	1. girl (5, 0.01%) 2. women (2, 0.01%) 3. woman (2, 0.01%) 4. womens (1) 5. girls (1) 6. ladies (1) 7. lady (1)
Time: New and young (T3--)	1. baby (32, 0.08%) 2. younger (4, 0.01%) 3. toddler (2)	1. younger (3, 0.01%) 2. baby (3, 0.01%) 3. toddler (1) 4. baby stuff (1) 5. toddlers (1)

2. Body and health		
Medicines and medical treatment (B3)	1. hospital (18, 0.04%) 2. balm (13, 0.03%) 3. treatment (10, 0.02%) 4. doctor (8. 0.02%) 5. ward (8, 0.02%) 6. hospitalized (8, 0.02%) 7. surgery (7, 0.02%) 8. massage (6, 0.01%) 9. dentist (6, 0.01%) 10. capsules (5, 0.01%)	1. dentist (10, 0.03%) 2. hospital (7, 0.02%) 3. extraction (7, 0.02%) 4. clinic (4, 0.01%) 5. doctor (3, 0.01%) 6. medical (3, 0.01%) 7. cardio (3, 0.01%) 8. Dr. (2, 0.01%) 9. treatment (2, 0.01%) 10. filling (2, 0.01%)
Anatomy and physiology (B1)	1. neck (19, 0.05%) 2. shoulder (17, 0.04%) 3. tired (16, 0.04%) 4. sleep (15, 0.04%) 5. eyes (15, 0.04%) 6. hair (12, 0.03%) 7. jaw (12, 0.03%) 8. heart (11, 0.03%) 9. body (10, 0.02%) 10. back (9, 0.02%)	1. sleep (12, 0.03%) 2. hands (10 ,0.03%) 3. body (9, 0.02%) 4. head (8, 0.02%) 5. legs (7, 0.02%) 6. face (6, 0.02%) 7. tired (5, 0.01%) 8. hand (5, 0.01%) 9. tooth (5, 0.01%) 10. physical (4, 0.01%)
Disease (B2-)	1. asthma (24, 0.06%) 2. pain (14,0.03%) 3. patients (7,0.02%) 4.patient (6, 0.01%) 5. disabled (6, 0.01%) 6. hurt (5, 0.01%) 7. ill (4, 0.01%) 8. symptoms (4, 0.01%) 9. painless (4, 0.01%) 10. madness (3, 0.01%)	1. fever (5, 0.01%) 2. ill (5, 0.01%) 3. dengue (4, 0.01%) 4.injuries (3, 0.01%) 5. injury (3, 0.01%) 6. sick (3, 0.01%) 7. painful (2, 0.01%) 8. cancer (2, 0.01%) 9. paralysis (2, 0.01%) 10. crazy (2, 0.01%)
3. Food		
Food (F1)	1. eat (30, 0.07%) 2. food (26, 0.06%) 3. dinner (21, 0.05%) 4. cook (21, 0.05%) 5. recipe (21, 0.05%) 6. breakfast (18, 0.04%) 7. rice (17, 0.04%) 8. cooking (16, 0.04%) 9. lunch (15, 0.04%) 10. restaurant (14, 0.03%)	1. food (36, 0.09%) 2. lunch (15, 0.04%) 3. restaurant (13, 0.03%) 4. dinner (12, 0.03%) 5. eat (12, 0.03%) 6. meals (9, 0.02%) 7. eating (9, 0.02%) 8. meal (8, 0.02%) 9. rice (8, 0.02%) 10. meat (8, 0.02%)

4. Others		
Parts of buildings (H2)	1. room (20, 0.05%) 2. department (7, 0.02%) 3. floor (6, 0.01%) 4. bedroom (4, 0.01%) 5. door (4, 0.01%) 6. wall (3, 0.01%) 7. walls (3, 0.01%) 8. hall (2) 9. suite (2) 10. living room (2)	1. door (6, 0.02%) 2. hall (5, 0.01%) 3. room (4, 0.01%) 4. reception (3, 0.01%) 5. bedroom (2, 0.01%) 6. wall (1) 7.locker room (1) 8. aisle (1) 9. department (1) 10. ballroom (1)
Arts and crafts (C1)	1. photo (26, 0.06%) 2. pictures (14, 0.03%) 3. photos (14, 0.03%) 4. art (14, 0.03%) 5. camera (10, 0.02%) 6. picture (9, 0.02%) 7. painting (7, 0.02%) 8. design (5, 0.01%) 9. designed (5, 0.01%) 10. paintings (4, 0.01%)	1. photos (8, 0.02%) 2. picture (6, 0.02%) 3. pictures (5, 0.01%) 4. camera (5, 0.01%) 5. photo (4, 0.01%) 6. cultural (4, 0.01%) 7. cultures (4, 0.01%) 8. designed (4, 0.01%) 9. paint (3, 0.01%) 10. culture (2, 0.01%)
Living creatures: animals, birds, etc. (L2)	1. chicken (29, 0.07%) 2. crabs (7, 0.02%) 3. dog (6, 0.01%) 4. eggs (6, 0.01%) 5. keeper (6, 0.01%) 6. crab (5, 0.01%) 7. fish (4, 0.01%) 8. bunny (4, 0.01%) 9. animals (4, 0.01%) 10. egg (4, 0.01%)	1. chicken (8, 0.02%) 2. cobra (5, 0.01%) 3. dog (5, 0.01%) 4. fish (3, 0.01%) 5. mussels (2, 0.01%) 6. cows (2, 0.01%) 7. turtle (2, 0.01%) 8. snake (2, 0.01%) 9. animal (2, 0.01%) 10. animals (2, 0.01%)
Shape (O4.4)	1. line (9, 0.02%) 2. angle (6, 0.01%) 3. cups (4, 0.01%) 4. corner (3, 0.01%) 5. flat (3, 0.01%) 6. loop (3, 0.01%) 7. 3D (3, 0.01%) 8. radial (2) 9. spot (2) 10. sharp (2)	1. shape (4, 0.01%) 2. square (3, 0.01%) 3. 3D (2, 0.01%) 4. sharp (2, 0.01%) 5. cross (2, 0.01%) 6. nuggets (1) 7. shaping (1) 8. line (1) 9. flat (1) 10. circle (1)

Getting and possession (A9+)	1. have (152, 0.37%) 2. get (81, 0.19%) 3. had (73, 0.18%) 4. has (49, 0.12%) 5. take (33, 0.08%) 6. keep (28, 0.07%) 7. got (26, 0.06%) 8. having (20, 0.05%) 9. took (16, 0.04%) 20. getting (12, 0.03%)	1. have (117, 0.30%) 2. get (47, 0.12%) 3. had (45, 0.11%) 4. has (31, 0.08%) 5. take (29, 0.07%) 6. got (28, 0.07%) 7. took (20, 0.05%) 8. keep (14, 0.04%) 9. having (13, 0.03%) 10. taking (11, 0.03%)

Key semantic domains in the Malaysian male weblogs: (Section 4.3.2)

Key semantic domain	Male weblogs	Female weblogs
1. Geography and Location		
Geographical names (Z2)	1. Chinese (29, 0.07%) 2. Malaysia (28, 0.07%) 3. Hanoi (20, 0.05%) 4. Korean (15, 0.04%) 5. Myanmar (12, 0.03%) 6. Cantonese (12, 0.03%) 7. English (11, 0.03%) 8. New York (9, 0.02%) 9. Singapore (8, 0.02%) 10. KL (8, 0.02%)	1. Malaysia (17, 0.04%) 2. Chinese (17, 0.04%) 3. KL (11, 0.03%) 4. China (9, 0.02%) 5. Penang (7, 0.02%) 6. Singapore (6, 0.01%) 7. Las Vegas (6, 0.01%) 8. Ipoh (5, 0.01%) 9. Istanbul (5, 0.01%) 10. city palace (4, 0.01%)
Geographical terms (W3)	1. island (16, 0.04%) 2. beach (6, 0.02%) 3. sea (4, 0.01%) 4. land (4, 0.01%) 5. rainforest (4, 0.01%) 6. atmosphere (3, 0.01%) 7. coast (3, 0.01%) 8. valley (3, 0.01%) 9. wave (3, 0.01%) 10. Mount (3, 0.01%)	1. lake (7, 0.02%) 2. beach (3, 0.01%) 3. sea (3, 0.01%) 4. pool (3, 0.01%) 5. earth (2) 6. hill (2) 7. slop (2) 8. cliffs (1) 9. tides (1) 10. coast (1)
Places (M7)	1. place (28, 0.07%) 2. city (18, 0.05%) 3. area (11, 0.03%) 4. town (10, 0.03%) 5. local (8, 0.02%) 6. international (8, 0.02%) 7. places (6, 0.02%) 8. countries (5, 0.01%) 9. village (4, 0.01%) 10. park (4, 0.01%)	1. place (18, 0.04%) 2. city (11, 0.03%) 3. places (10, 0.02%) 4. entry (8, 0.02%) 5. area (5, 0.01%) 6. countries (4, 0.01%) 7. town (3, 0.01%) 8. international (3, 0.01%) 9. hometown (3, 0.01%) 10. location (3, 0.01%)
2. Leisure		
Music and related activities (K2)	1. music (21, 0.05%) 2. hip-hop (15, 0.04%) 3. song (5, 0.01%) 4. sang (4, 0.01%) 5. songs (3, 0.01%) 6. singing (3, 0.01%) 7. album (3, 0.01%)	1. song (6, 0.01%) 2. musical (4, 0.01%) 3. music (4, 0.01%) 4. jazz (2) 5. tuned (2) 6.violin (1) 7. scrapbook (1)

	8. pop (2, 0.01%) 9. guitar (2, 0.01%) 10. tune (2, 0.01%)	8. chorine (1) 9. drum (1) 10. singing (1)
3. Money and Industry		
Money: Affluence (I1.1+)	1. prosperity (5, 0.01%) 2. wealth (5, 0.01%) 3. bonus (5, 0.01%) 4. prosperous (2, 0.01%) 5. wealthy (1) 6. affluent (1) 7. refund (1) 8. refunded (1) 9. well-to-do (1)	1. prosperous (1) 2. rich (1)
Industry (I4)	1. mine (5, 0.01%) 2. factory (3, 0.01%) 3. workshops (3, 0.01%) 4. mining industry (2, 0.01%) 5. mining (2, 0.01%) 6. works (2, 0.01%) 7. aviation industry (2, 0.01%) 8. pit mines (1) 9. IT industries (1) 10. industry (1)	1. workshops (1) 2. mine (1) 3. factory (1) 4. industry (1)
Money and pay (I1.1)	1. profit (8, 0.02%) 2. investment (5, 0.01%) 3. fund (5, 0.01%) 4. capital (4, 0.01%) 5. afford (3, 0.01%) 6. tax (3, 0.01%) 7. save (2, 0.01%) 8. earnings (2, 0.01%) 9. investments (2, 0.01%) 10. dividends (2, 0.01%)	1. save (10, 0.02%) 2. investment (4, 0.01%) 3. pay (2) 4. saving (1) 5. tax (1) 6. saved (1) 7. banking (1) 8. income (1) 9. afford (1) 10. savings (1)
4. Comparing and Evaluation		
Comparing: Similar/different (A6.1)	1. compared (8, 0.02%) 2. compare (6, 0.02%) 3. comparison (3, 0.01%) 4. comparisons (1)	1. compared (2) 2. balance out (1)
Comparing: Usual (A6.2+)	1. usually (10, 0.03%) 2. common (10, 0.03%)	1. tend (8, 0.02%) 2. normal (7, 0.02%)

	3. regular (10, 0.03%) 4. normal (6, 0.02%) 5. tend (5, 0.01%) 6. generally (5, 0.01%) 7. wont (4, 0.01%) 8. usual (4, 0.01%) 9. natural (4, 0.01%) 10. familiar (4, 0.01%)	3. common (7, 0.02%) 4. basic (6, 0.01%) 5. as usual (5, 0.01%) 6. usual (4, 0.01%) 7. wont (4, 0.01%) 8. usually (3, 0.01%) 9. naturally (3, 0.01%) 10, average (1)
Important (A11.1+)	1. important (17, 0.04%) 2. main (14, 0.04%) 3. value (5, 0.01%) 4. major (4, 0.01%) 5. matter (4, 0.01%) 6. serious (3, 0.01%) 7. values (3, 0.01%) 8. significant (3, 0.01%) 9. upgraded (2, 0.01%) 10. upgrade (2, 0.01%)	1. major (4, 0.01%) 2. important (3, 0.01%) 3. main (2) 4. urgent (2) 5. means a lot (1) 6. well known (1) 7. prominent (1) 8. emergency (1) 9. make all the difference (1) 10. biggies (1)
5. Measurement and Quantifiers		
Measurement: Weight (N3.5)	1. weight (11, 0.03%) 2. 130 kg (5, 0.01%) 3. pressures (3, 0.01%) 4. 5 kg (3, 0.01%) 5. 45 kg (2, 0.01%) 6. overweight (2, 0.01%) 7. 1 kg (2, 0.01%) 8. weighs (2, 0.01%) 9. 2 kg (1) 10. jin (1)	1. weight (2) 2. 21 b (1) 3. 2-31 b (1) 4. kilograms (1) 5. 1 bs (1) 6. pounder (1) 7. pressure (1) 8. heaviness (1)
Quantities: many/much (N5+)	1. much (33, 0.08%) 2. many (32, 0.08%) 3. enough (27, 0.07%) 4. a lot (14, 0.04%) 5. lots (10, 0.03%) 6. load (8, 0.02%) 7. increase (7, 0.02%) 8. plenty (6, 0.02%) 9. general (6, 0.02%) 10. rise (6, 0.02%)	1. much (29, 0.07%) 2. enough (24, 0.06%) 3. a lot (16, 0.04%) 4. many (11, 0.03%) 5. lots (9, 0.02%) 6. add (9, 0.02%) 7. bunch (8, 0.02%) 8. couple (3, 0.01%) 9. lotsa (3, 0.01) 10. increase (3, 0.01%)

6. Social Actions and Process		
Participating (S1.1.3+)	1. conference (12, 0.03%) 2. meeting (9, 0.02%) 3. reunion (6, 0.02%) 4. parties (2, 0.01%) 5. collaboration (2, 0.01%) 6. attending (2, 0.01%) 7. attended (2, 0.01%) 8. meetings (1) 9. forum (1) 10. participating (1)	1. meeting (2) 2. conference (2) 3. attend (2) 4. participation (1) 5. reunion (1) 6. meetings (1) 7. participating (1) 8. met up (1) 9. joined in (1) 10. conference room (1)
Competition (S7.3)	1. contest (3, 0.01%) 2. finalists (2, 0.01%) 3. tournament (1) 4. tournaments (1) 5. opponent (1) 6. rivalries (1)	
Religion and the supernatural (S9)	1. God (19, 0.05%) 2. soul (11, 0.03%) 3. church (8, 0.02%) 4. Christmas (6, 0.02%) 5. holy spirit (5, 0.01%) 6. spirit (5, 0.01%) 7. Tao (5, 0.01%) 8. temples (4, 0.01%) 9. holy (3, 0.01%) 10. Islam (3, 0.01%)	1. Christmas (10, 0.02%) 2. God (7, 0.02%) 3. witch (5, 0.01%) 4. soul (4, 0.01%) 5. pray (4, 0.01%) 6. hell (3, 0.01%) 7. heaven (3, 0.01%) 8. sacrificing (3, 0.01%) 9. temples (2) 10. Muslim (2)
7. Others		
Warfare, defence and the army; weapons (G3)	1. army (9, 0.02%) 2. marines (7, 0.02%) 3. guns (5, 0.01%) 4. shot (3, 0.01%) 5. WWII (3, 0.01%) 6. Crossbow (2, 0.01%) 7. shoot (2, 0.01%) 8. shooting (2, 0.01%) 9. guardsmen (2, 0.01%) 10. armour (2, 0.01%)	1. shot (5, 0.01%) 2. firework (4, 0.01%) 3. shooting (4, 0.01%) 4. bullet (3, 0.01%) 5. gun (2) 6. arrow (1) 7. guns (1) 8. invasion (1) 9. breech (1) 10. shooting away (1)
Grammatical bin (Z5)	1. the (1859, 4.74%) 2. and (1123, 2.86%) 3. to (1103, 2.81%) 4. a (777, 1.98%)	1. the (1623, 3.90%) 2. to (1212, 2.91%) 3. and (1155, 2.78%) 4. a (794, 1.91%)

	5. of (758, 1.93%) 6. in (518, 1.32%) 7. for (423, 1.08%) 8. with (275, 0.70%) 9. but (232, 0.59%) 10. on (196, 0.50%)	5. of (572, 1.38%) 6. for (421, 1.01%) 7. in (397, 0.95%) 8. with (288, 0.69%) 9. but (260, 0.63%) 10. on (186, 0.45%)
Information technology and computing (Y2)	1. blog (38, 0.10%) 2. blogger (10, 0.03%) 3. website (9, 0.02%) 4. digital (7, 0.02%) 5. IT (7, 0.02%) 6. blogs (7, 0.02%) 7. Internet (5, 0.01%) 8. blogging ((4, 0.01%) 9. bloggers (4, 0.01%) 10. websites (4, 0.01%)	1. blog (40, 0.10%) 2. online (5, 0.01%) 3. bloggers (5, 0.01%) 4. Internet (4, 0.01%) 5. blogging (3, 0.01%) 6. screen (3, 0.01%) 7. laptop (2) 8. screens (2) 9. blogged (2) 10. blogs (2)
Existing (A3+)	1. is (487, 1.24%) 2. was (216, 0.55%) 3. be (181, 0.46%) 4. are (171, 0.44%) 5. ’s (115, 0.29%) 6. were (39, 0.10%) 7. am (29, 0.07%) 8. been (26, 0.07%) 9. ’m (26, 0.07%) 10. being (26, 0.07%)	1. is (365, 0.88%) 2. was (254, 0.61%) 3. be (148, 0.36%) 4. are (127, 0.31%) 5. ’s (116, 0.28%) 6. am (54, 0.13%) 7. ’m (44, 0.11%) 8. been (41, 0.10%) 9. were (40, 0.10%) 10. being (35, 0.08%)
Expected (X2.6+)	1. hope (33, 0.08%) 2. expect (9, 0.02%) 3. hopefully (6, 0.02%) 4. looking forward (5, 0.01%) 5. hoping (5, 0.01%) 6. expected (5, 0.01%) 7. hopes (4, 0.01%) 8. look forward (2, 0.01%) 9. expectations (2, 0.01%) 10. expecting (2, 0.01%)	1. hope (15, 0.04%) 2. expect (7, 0.02%) 3. expected (6, 0.01%) 4. looking forward (4, 0.01%) 5. expectation (3, 0.01%) 6. hopefully (3, 0.01%) 7. hopes (1) 8. anticipated (1) 9. due (1) 10. foresee (1)

Appendix 5: Wmatrix Key Parts of Speech (POS) in the Corpora (Chapter 4)

Key parts of speech in the Malaysian female weblogs: (Section 4.4.1)

Key parts of speech	Female weblogs	Male weblogs
1. Personal Pronouns		
3rd person sing. subjective personal pronoun (*he, she*) (PPHS1)	1. he (166, 0.40%) 2. she (121, 0.29%)	1. he (94, 0.24%) 2. she (19, 0.05%)
3rd person sing. objective personal pronoun (*him, her*) (PPHO1)	1. him (67, 0.16%) 2. her (50, 0.12%)	1. him (23, 0.06%) 2. her (15, 0.04%)
1st person sing. subjective personal pronoun (*I*) (PPIS1)	1. I (1291, 3.10%) 2. I think (13, 0.03%) 3. I mean (5, 0.01%) 4. I guess (4, 0.01%) 5. I say (2)	1. I (1019, 2.60%) 2. I think (6, 0.02%) 3. I see (4, 0.01%) 4. I mean (3, 0.01%) 5. I guess (3, 0.01%) 6. I say (3, 0.01%) 7. I suppose (1) 8. I have to say (1) 9. I bet (1)
2. Lexical Verbs		
past tense of lexical verb (VVD)	1. said (45, 0.11%) 2. went (37, 0.09%) 3. got (36, 0.09%) 4. thought (25, 0.06%) 5. asked (25, 0.06%) 6. decided (23, 0.06%) 7. told (21, 0.05%) 8. made (21, 0.05%) 9. started (18, 0.04%) 10. took (16, 0.04%)	1. got (29, 0.07%) 2. felt (20, 0.05%) 3. took (20, 0.05%) 4. said (18, 0.05%) 5. made (16, 0.04%) 6. bought (16, 0.04%) 7. went (15, 0.04%) 8. thought (15, 0.04%) 9. decided (13, 0.03%) 10. wanted (13, 0.03%)
base form of lexical verb (VV0)	1. know (45, 0.11%) 2. need (42, 0.10%) 3. love (40, 0.10%) 4. want (37, 0.09%) 5. get (35, 0.08%) 6. think (29, 0.07%) 7. thank you (26, 0.06%) 8. let (21, 0.05%)	1. know (38, 0.10%) 2. think (28, 0.07%) 3. thank you (27, 0.07%) 4. want (26, 0.07%) 4. hope (26, 0.07%) 6. need (23, 0.06%) 7. remember (21, 0.05%) 8. I'm (19, 0.05%)

	9. like (21, 0.05%) 10. guess (18, 0.04%)	9. get (19, 0.05%) 10. like (14, 0.04%)
3. Interjection		
interjection (UH)	1. yes (22, 0.05%) 2. haha (18, 0.04%) 3. oh (18, 0.04%) 4. yeah (12, 0.03%) 5. hello (11, 0.03%) 6. no (10, 0.02%) 7. ah (9, 0.02%) 8. wow (8, 0.02%) 9. huh (6, 0.01%) 10. oh well (6, 0.01%)	1. yes (17, 0.04%) 2. yeah (8, 0.02%) 3. oh (7, 0.02%) 4. no (6, 0.02%) 5. haha (5, 0.01%) 6. hello (5, 0.01%) 7. ah (3, 0.01%) 8. alhamdulilla (2, 0.01%) 9. ha (2, 0.01%) 10. hey (2, 0.01%)
4. General Adverb		
general adverb (RR)	1. just (148, 0.36%) 2. so (127, 0.31%) 3. really (79, 0.19%) 4. also (65, 0.16%) 5. still (65, 0.16%) 6. only (62, 0.15%) 7. well (59, 0.14%) 8. always (45, 0.11%) 9. even (45, 0.11%) 10. too (42, 0.10%)	1. just (111, 0.28%) 2. so (76, 0.19%) 3. also (68, 0.17%) 4. really (67, 0.17%) 5. well (48, 0.12%) 6. still (44, 0.11%) 7. only (42, 0.11%) 8. always (37, 0.09%) 9. even (35, 0.09%) 10. never (28, 0.07%)
5. Copular verb *be*		
am (VBM)	1. am (83, 0.20%) 2. ’m (82, 0.20%) 3. am about to (1)	1. ’m (49, 0.12%) 2. am (48, 0.12%) 3. am back (3, 0.01%) 4. am about to (1) 5. ’m back (1)

Key parts of speech in the Malaysian male weblogs: (Section 4.4.2)

Key parts of speech	Male weblogs	Female weblogs
1. Prepositions		
of (as preposition) (IO)	1. of (751, 1.91%) 2. of the day (1)	1. of (570, 1.37%) 2. of the day (1) 3. of a time (1)
general preposition (II)	1. in (518, 1.32%) 2. to (325, 0.83%) 3. on (196, 0.50%) 4. from (149, 0.38%) 5. at (141, 0.36%) 6. by (106, 0.27) 7. about (94, 0.24%) 8. like (79, 0.20%) 9. as (53, 0.14%) 10. after (44, 0.11%)	1. in (397, 0.95%) 2. to (345, 0.83%) 3. on (186, 0.45%) 4. at (176, 0.42%) 5. from (155, 0.37%) 6. by (97, 0.23%) 7. about (88, 0.21%) 8. like (73, 0.18%) 9. after (53, 0.13%) 10. before (48, 0.12%)
2. Article		
article (AT)	1. the (1859, 4.74%) 2. no (57, 0.15%) 3. the rainbow (3, 0.01%) 4. the bill (3, 0.01%) 5. - the (3, 0.01%) 6. the birds (2, 0.01%) 7. the year before (2, 0.01%) 8. the sun (2, 0.01%) 9. the olden days (1) 10. the likeness of (1)	1. the (1622, 3.90%) 2. no (49, 0.12%) 3. the night before (2) 4. the royal (2) 5. the bill (2) 6. the stairs (2) 7. the other night (1) 8. the hell (1) 9. the ocean (1) 10. the sun (1)
3. General Adjective		
general adjective (JJ)	1. good (67, 0.17%) 2. new (42, 0.11%) 3. happy (40, 0.10%) 4. big (35, 0.09%) 5. other (32, 0.08%) 6. Chinese (29, 0.07%) 7. great (29, 0.07%) 8. new year (27, 0.07%) 9. old (24, 0.06%) 10. beautiful (23, 0.06%)	1. good (64, 0.15%) 2. new (53, 0.13%) 3. happy (50, 0.12%) 4. little (37, 0.09%) 5. big (36, 0.09%) 6. old (35, 0.08%) 7. other (25, 0.06%) 8. new year (23, 0.06%) 9. whole (22, 0.05%) 10. bad (21, 0.05%)

4. Singular Proper Noun		
singular proper noun (NP1)	1. Malaysia (28, 0.07%) 2. Hanoi (20, 0.05%) 3. God (17, 0.04%) 4. Altantuya (13, 0.03%) 5. Myanmar (12, 0.03%) 6. PM (10, 0.03%) 7. Seoul (10, 0.03%) 8. New York (9, 0.02%) 9. Singapore (8, 0.02%) 10. Gangnam (7, 0.02%)	1. Malaysia (17, 0.04%) 2. Henry (12, 0.03%) 3. KL (11, 0.03%) 4. Bukit Jugra (8, 0.02%) 5. Peter (8, 0.02%) 6. China (8, 0.02%) 7. God (7, 0.02%) 8. Penang (7, 0.02%) 9. Ethan (7, 0.02%) 10. Jaipur (6, 0.01%)
5. Copular verb *be*		
is (VBM)	1. is (602, 1.53%) 2. 's (128, 0.33%) 3. is after (2, 0.01%) 4. is over (1) 5. is out to (1) 6. 's over (1) 7. is up (1) 8. is yourself (1) 9. is the case (1)	1. is (449, 1.08%) 2. 's (128, 0.31%) 3. is against (1) 4. is after (1) 5. 's over (1) 6. is over (1)
are (VBR)	1. are (236, 0.60%) 2. 're (11, 0.03%) 3. are ourselves (1)	1. are (162, 0.39%) 2. 're (11, 0.03%)
6. Formula		
formula (FO)	1. = (8, 0.02%) 2. AT 890 (7, 0.02%) 3. GE 13 (4, 0.01%) 4. Dre1M (4, 0.01%) 5. Daai 6 (3, 0.01%) 6. TX2032 (2, 0.01%) 7. TV3 (2, 0.01%) 8. 5 kgs (2, 0.01%) 9. C4 (2, 0.01%) 10. RM 45 (2, 0.01%)	1. + (9, 0.02%) 2. RVD+VE (3, 0.01%) 3. = (3, 0.01%) 4. RM 300 (3, 0.01%) 5. RM 500 (3, 0.01%) 6. RM 9 (2) 7. / (2) 8. RM 60 (2) 9. RM 5 (2) 10. RM 50 (2)