A Corpus-Based Analysis of Differences in Language Use by Men and Women Bloggers of English E-Newspapers in Pakistan

Researcher

Ibrar Hussain Khan

Supervisor

Dr. Ayaz Afsar

Reg. No. 80-FLL/PHDENG/F14

Department of English

Faculty of Languages and Literature

INTERNATIONAL ISLAMIC UNIVERSITY, ISLAMABAD

2019

i Language and languages — sex differences

ii Language and sex

iii Newspaper

iv Women

A Corpus-Based Analysis of Differences in Language Use by Men and Women Bloggers of English E-Newspapers in Pakistan

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy, at the Department of English, Faculty of Languages and Literature, International Islamic University, Islamabad

Ibrar Hussain Khan

Reg. No. 80-FLL/PHDENG/F14

Department of English
Faculty of Languages and Literature
International Islamic University, Islamabad
2019

Acceptance by the *Viva Voce* Committee

Title of the thesis: **A Corpus-Based Analysis of Differences in Language Use by Men and Women Bloggers of English E-Newspapers in Pakistan**

Name of Student: **Ibrar Hussain Khan**

Registration No: **80-FLL/PHDENG/F14**

Accepted by the department of English, Faculty of Languages & Literature, International Islamic University, Islamabad, in partial fulfillment of the requirement for the Doctor of Philosophy degree in English.

Viva Voce Committee

External Examiner-1
Prof. Dr. Mobina Talaat
Professor Linguistics, House No. 1533,
Cornice Road, Phase-4 Bahria town,
Islamabad

Prof. Dr. Ayaz Afsar
Dean
Faculty of Languages and Literature,
IIUI

External Examiner-2
Dr. G. Raza
Associate Professor, PIEAS, Islamabad

Dr. Muhammad Sheeraz
Incharge
Department of English, FLL

Internal Examiner
Dr. Akhtar Aziz
Assistant Professor/HOD,
Center of Excellence for Modern
Languages (CEML)
International Islamic University
Islamabad, IIUI

Supervisor
Prof. Dr. Ayaz Afsar
Dean
Faculty of Languages and Literature,
IIUI

December 6, 2019

DECLARATION

I, Ibrar Hussain Khan, Registration No. 80-FLL/PHDENG/F14, a student of PhD (English) at the International Islamic University, Islamabad, hereby declare that, during the period of this study, I was not registered in any other degree programme and that this thesis has not been submitted for any degree or other purposes. The material produced in this thesis titled "A Corpus-Based Analysis of Differences in Language Use by Men and Women Bloggers of English E-Newspapers in Pakistan" has not been submitted by me wholly or in part for any other academic award or qualification and shall not be submitted by me in future for obtaining any degree from this or any other university. I confirm that this thesis is the original work of the researcher except where otherwise acknowledged in the thesis. I also understand that if evidence of plagiarism is found in my thesis/dissertation at any stage, even after the award of a degree, the work may be cancelled and the degree revoked.

(Ibrar Hussain Khan)

يَٰٓأَيُّهَا ٱلنَّاسُ إِنَّا خَلَقْنَٰكُم مِّن ذَكَرٍ وَأُنثَىٰ وَجَعَلْنَٰكُمْ شُعُوبًا وَقَبَآئِلَ لِتَعَارَفُوٓا۟ ۚ إِنَّ أَكْرَمَكُمْ
عِندَ ٱللَّهِ أَتْقَىٰكُمْ ۚ إِنَّ ٱللَّهَ عَلِيمٌ خَبِيرٌ ﴿١٣﴾

O mankind, We have created you from a male and a female, and made you into races and tribes, so that you may identify one another. Surely the noblest of you, in Allah's sight, is the one who is most pious of you. Surely Allah is All-Knowing, All-Aware.

—*Al-Quran* (49:13)
Translation by Mufti Taqi Usmani

وَمِنْ ءَايَٰتِهِۦ خَلْقُ ٱلسَّمَٰوَٰتِ وَٱلْأَرْضِ وَٱخْتِلَٰفُ أَلْسِنَتِكُمْ وَأَلْوَٰنِكُمْ ۚ إِنَّ فِى ذَٰلِكَ
لَءَايَٰتٍ لِّلْعَٰلِمِينَ ﴿٢٢﴾

And among His Signs is the creation of the heavens and the earth, and the variations in your languages and your colours: verily in that are Signs for those who know.

—*Al-Quran* (30:22)
Translation by Abdullah Yusuf Ali

ACKNOWLEDGEMENTS

All praise be to Allah Almighty who gave me the strength and ability to complete this work. Certainly, it is He alone who created mankind from a man and a woman and declared language variations among human beings as one of His Signs. And, may Allah Almighty's countless blessings and infinite peace be upon His final and most-beloved Messenger sent for the entire mankind—Muhammadﷺ—who abolished all ethnic and linguistic differences as symbols of superiority or power in human beings and made for them piety as the point of departure; who said that he who does not thank the people is not thankful to Allah Almighty (Sunan Abi Dawud: 4811).

Indeed, the present work would not have been completed without the personal attention and consistent support of my research supervisor Professor, Dr. Ayaz Afsar. The guidance he provided, while delivering comprehensive lectures on corpus linguistics and statistics as well as conducting methodical sessions on application of relevant text analysis tools during teaching us the module of *Corpora in Applied Linguistics*, was of immense value for me to muster enough motivation and confidence to undertake the present work. He was also instrumental in arranging an exclusive international workshop on corpus linguistics at International Islamic University Islamabad (IIUI), in collaboration with the University of Birmingham (UoB), the University of Leeds and the British Council, Pakistan, which provided me an opportunity to interact with international academia, get exposure to a variety of corpus linguistics research methods and gain more hands-on-job training on text analysis tools from experts. He also guided, encouraged and helped me during the entire process of International Research Initiative Programme (IRSIP) scholarship offered by Higher Education Commission (HEC) of Pakistan.

I also feel indebted to HEC for the award of IRSIP scholarship, which afforded me a six-month stay as a Visiting Research Scholar at Centre for Advanced Research in English (CARE), UoB. In this regard, I am also extremely grateful to Dr. Ameer Sultan Awan of IIUI who, prior to my departure, provided useful information at each stage of my IRSIP scholarship award, visa process and lodging arrangements in the UK.

At CARE, the thorough guidance by Dr. Paul Thompson, Deputy Director of the Centre for Corpus Research (my academic advisor) on methodological and theoretical aspects of my study as well as the directions given by Dr. Bodo Winter, Lecturer in Cognitive Linguistics, on statistical procedures and tests conducted for my study were of immense help. Special credit also goes to Dr. Paul Thompson, Dr. Florent Perek and Viola Wiegand for conducting corpus tools training sessions. I also learnt a lot from various faculty members and students who presented their works in *English Language Research Staff Seminars* and *PhD Student Research Seminars* respectively. Also, the regular classes on Academic Writing, Academic Study Skills, Communication and Presentation Skills and Advanced English Grammar, arranged by the British International Academy (BIA) of UoB, greatly helped me in writing this thesis. I am also grateful to Dr. Suganthi John, Director CARE, and Michelle Devereux, Administrator CARE, for facilitating my visit before and after my arrival at CARE.

I am also indebted to my first cousin Muhammad Mehtab, a software engineer by profession, who developed the custom-built software "FileAttributeReader" used in this study. A note of sincere thanks is also due to my friend and colleague Mr. Muhammad Amjad, who took the pains to proofread the drafts. Last but not the least, my entire family members deserve special credit for their constant patience, support, encouragement, and prayers during the entire course of my study, both inland and abroad.

TABLE OF CONTENTS

Acknowledgments vii

Abbreviations and Symbols xvii

List of Tables xviii

List of Figures xx

List of Concordances xxiii

Abstract xxv

CHAPTER 1: INTRODUCTION

1.1 Chapter Overview 1

1.2 Motivation for the Study 1

1.3 Background to the Study 2

1.4 Statement of Problem 2

1.5 Hypotheses 3

1.6 Justification of the Topic 3

1.7 Delimitation of the Study 5

1.8 Significance of the Study 7

1.9 Outline of Chapters 9

CHAPTER 2: LITERATURE REVIEW

2.1 Chapter Overview 11

2.2 Theory: Language and Gender 11

2.2.1 The Field Takes a New Shape 13

2.2.2 Theorising Language and Gender 14

CHAPTER 5: DISCUSSIONS

5.1 Chapter Overview..... 93

5.2 Word Count 93

5.3 Summary Language Features 95

5.3.1 Analytical Thinking 95

5.3.2 Clout 96

5.3.3 Authentic/Dictionary Words 98

5.3.4 Emotional Tone 99

5.3.5 Words Per Sentence 101

5.3.6 Words > 6 Letters 102

5.4 Linguistics Dimensions 104

5.4.1 Function Words 104

5.4.2 Total Pronouns 106

5.4.3 Personal Pronouns 108

5.4.4 First Person Singular 108

5.4.5 First Person Plural 114

5.4.6 Second Person Pronoun 118

5.4.7 Third Person Singular Pronoun (*he/she*) 123

5.4.8 Third Person Plural Pronoun (*they*) 131

5.4.9 Impersonal Pronouns 139

5.4.10 Articles..... 141

5.4.11 Prepositions 142

5.4.12 Auxiliary Verbs and Common Verbs 145

Money 216

Religion 216

Death 219

5.7 Informal Language 220

5.7.1 Swear Words..... 221

5.7.2 Netspeak 222

5.7.3 Assent..... 224

5.7.4 Nonfluencies 224

5.7.5 Fillers 225

5.8 Total Punctuation 226

5.8.1 Commas..... 228

5.8.2 Semicolons..... 228

5.8.3 Question Marks 229

5.8.4 Exclamation Marks 230

5.8.5 Dashes 231

5.8.6 Quotation Marks 231

5.8.7 Apostrophes..... 232

CHAPTER 6: CONCLUSIONS

6.1 Overview of the Study 233

6.2 Key Findings 234

6.2.1 Methodological Aspects 234

6.2.2 Linguistic Aspects..... 235

6.2.3 Gender Effect on Language 238

6.3 Additional Findings 240

6.3.1 Grammatical & Lexical Aspects 240

6.3.2 Social Aspects 242

6.3.3 Psychological Aspects 243

6.3.4 Biological Aspects 244

6.3.5 Routines 244

6.4 Strengths and Limitations 245

6.5 Recommendations for Future Research 246

References 248

Appendix A LIWC2015 Output Variable Information 296

Appendix B Sample of LIWC2015 Results 299

Appendix C Mann-Whitney U-Test Results 300

Abbreviations

- ATA Automated Text Analysis
- BBC British Broadcasting Corporation
- BNC British National Corpus
- CL Corpus Linguistics
- CMC Computer-Mediated Communication
- CoP Community of Practice
- D Dawn
- EFL English as a Foreign Language
- GenCorB Gendered Corpus of Bloggers; a term used in the study where both MenCorB and WenCorB are meant together.
- JB Jarque-Berra
- LIWC Linguistic Inquiry and Word Count
- MenCorB Men Corpus of Bloggers
- MP Member of Parliament
- ns Not Significant
- POS Part-of-Speech
- PP Probability-Probability
- QQ Quantile-Quantile
- SD Standard Deviation
- SMS Short Message Service
- SBE Standard British English
- TDT The Daily Times
- TExT The Express Tribune
- TNn The Nation
- TNw The News
- WaC Web as Corpus
- WenCorB Women Corpus of Bloggers
- WfC Web for Corpus
- WFP Word Formation Process
- WWW World Wide Web

Symbols

- α : alpha
- H_1 : alternate hypothesis
- d : effect size
- $>$: greater than
- $<$: less than
- $\bar{X}$: mean
- H_0 : null hypothesis
- σ : standard deviation

List of Tables

Table 2.1	Sex ratio: Population in Pakistan	41
Table 2.2	Outline introduction of English e-newspapers chosen for the study	49
Table 3.1	Typology of GenCorB	54
Table 3.2	e-Newspaper-wise dates of oldest blog posts	57
Table 3.3	Different categories of bloggers' professions in the sample: Alphabetical order	65
Table 4.1	Results of Jarque-Berra normality test for all LIWC2015 language dimensions	78
Table 4.2	Hypotheses tests and main effects of gender on language	89
Table 5.1	Prepositions used more by men (frequencies normalised to one million	143
Table 5.2	Prepositions used more by women (frequencies normalised to one million)	144
Table 5.3	Results of adverb usage compared to previous research	148
Table 5.4	Comparison of results for selected adjectives (frequencies normalised to one million)	156
Table 5.5	Comparison of top 20 positive emotion words (frequencies normalised to one million)	163
Table 5.6	Comparison of top 20 negative emotion words (frequencies normalised to one million)	164
Table 5.7	List of words related to body parts used more by men (frequencies normalised to one million)	187
Table 5.8	List of words related to body parts used more by women (frequencies normalised to one million)	187
Table 5.9	Specific nouns used by women for expressing '*affiliation*' (frequencies normalised to one million)	197
Table 5.10	Specific nouns used by men for expressing '*affiliation*' (frequencies normalised to one million)	198

Table 5.11 Gender differences in themes related to '*leisure*' 213

Table 5.12 Gender differences in themes related to '*home*' 215

Table 5.13 Gender differences in themes related to '*religion*' 218

List of Figures

Figure 3.1	Percentage of bloggers' professions in GenCorB	65
Figure 4.1	Screenshot showing sample LIWC2015 results display. The column of file names has been squeezed for this screenshot to ensure bloggers' anonymity.	74
Figure 4.2	Screenshot of sample LIWC2015 results exported to MS Excel file. The column of file names has not been included in this screenshot to ensure bloggers' anonymity.	75
Figure 4.3	Histogram showing the non-normality of sampled data from MenCorB for LIWC2015 variable 'Pronoun'	83
Figure 4.4	Histogram showing the non-normality of sampled data from WenCorB for LIWC2015 variable 'Pronoun'	84
Figure 4.5	PP plot showing the non-normality of sampled data from MenCorB for LIWC2015 variable 'Pronoun'	84
Figure 4.6	PP plot showing the non-normality of sampled data from WenCorB for the LIWC2015 variable 'Pronoun'	85
Figure 4.7	QQ plot showing the non-normality of sampled data from MenCorB for LIWC2015 variable 'Pronoun'	85
Figure 4.8	QQ plot showing the non-normality of sampled data from WenCorB for LIWC2015 variable 'Pronoun'	86
Figure 5.1	Comparison of first person singular pronoun (frequencies normalised to one million)	109
Figure 5.2	Screenshot of "Methinks" frequency over the years: Source Collins online dictionary	110
Figure 5.3	Noun referents (masculine, feminine, neuter) to family members collocating with 'My' in GenCorB (frequencies normalised to one million)	112
Figure 5.4	First person plural pronoun (frequencies normalised to one million)	115
Figure 5.5	Second person pronoun variants (frequencies normalized to one million)	120

Figure 5.6	Personal third person singular pronoun variants (frequencies normalised to one million)	126
Figure 5.7	R1 noun-referents (masculine, feminine, neuter) to family members with 'His' in GenCorB (frequencies normalised to one million)	126
Figure 5.8	R1 noun-referents (masculine, feminine, neuter) to family members with 'Her' in GenCorB (frequencies normalised to one million)	127
Figure 5.9	Comparison of third person plural pronoun variants	134
Figure 5.10	'Their' collocations with nouns of family members (frequencies normalised to one million)	134
Figure 5.11	Contracted forms: Impersonal Pronoun Variants (frequencies normalised to one million)	140
Figure 5.12	Adverbs (frequencies normalised to one million)	147
Figure 5.13	Conjunctions (frequencies normalised to one million)	150
Figure 5.14	Negations used more by men (frequencies normalised to one million)	152
Figure 5.15	Negations used more by women (frequencies normalised to one million)	152
Figure 5.16	Colour adjectives based on Berlin and Kay's (1969) colour taxonomy (frequencies normalised to one million)	155
Figure 5.17	Negative emotion words used more by men (frequencies normalised to one million)	165
Figure 5.18	Negative emotion words used more by women (frequencies normalised to one million)	165
Figure 5.19	Anxiety words used more by women (frequencies normalised to one million)	167
Figure 5.20	Anxiety words used more by men (frequencies normalised to one million)	167
Figure 5.21	Words related to the theme of 'marriage' used more by women (frequencies normalised to one million)	172

Figure 5.22	Words related to 'friends' used more by men (frequencies normalised to one million)	172
Figure 5.23	Certainty words used more by men (frequencies normalised to one million)	180
Figure 5.24	Certainty words used more by women (frequencies normalised to one million)	180
Figure 5.25	Words related to colours; frequencies normalized to one million	183
Figure 5.26	Words related to the things seen (frequencies normalised to one million)	183
Figure 5.27	Words related to rape, abortion and prostitution (frequencies normalised to one million)	192
Figure 5.28	Words related to 'ingestion' overused by men (frequencies normalised to one million)	194
Figure 5.29	Words related to 'work' used more in MenCorB (frequencies normalised to one million)	210
Figure 5.30	Words related to 'work' used more in WenCorB (frequencies normalised to one million)	211
Figure 6.1	Summary of Effect Sizes	239

List of Concordances

Concordance 5.1	'Methinks': MenCorB	111
Concordance 5.2	'Methinks': WenCorB	111
Concordance 5.3	Spokesperson of 'nation': MenCorB	116
Concordance 5.4	Spokesperson of 'nation': WenCorB	116
Concordance 5.5	Spokesperson of 'ordinary citizen': MenCorB	116
Concordance 5.6	Spokesperson of 'ordinary citizen': WenCorB	117
Concordance 5.7	Spokesperson of 'cricket team': MenCorB	117
Concordance 5.8	Spokesperson of 'same-sex group': WenCorB	118
Concordance 5.9	*you* the sense of the 'addressee': MenCorB	121
Concordance 5.10	*you* in the sense of the 'addressee': WenCorB	121
Concordance 5.11	*you* in the sense of 'everybody': MenCorB	122
Concordance 5.12	*you* in the sense of 'everybody': WenCorB	122
Concordance 5.13	*you* as gender-directed 'everybody': MenCorB	122
Concordance 5.14	you as gender-directed 'everybody': WenCorB	123
Concordance 5.15	'*his* grandfather' occurrences in GenCorB	129
Concordance 5.16	'*her* grandfather' occurrences in GenCorB	129
Concordance 5.17	'*his* grandmother' occurrences in GenCorB	130
Concordance 5.18	'*her* grandmother' occurrences in GenCorB	131
Concordance 5.19	Affiliative context of 'they': MenCorB	136
Concordance 5.20	Affiliative context of 'they': WenCorB	137
Concordance 5.21	Oppositional context of 'they': MenCorB	137
Concordance 5.22	Oppositional context of 'they': WenCorB	138
Concordance 6.1	*you* as gender-directed 'everybody': MenCorB	240

Concordance 6.2	*you* as gender-directed 'everybody': WenCorB	240
Concordance 6.3	Spokesperson of 'cricket team': MenCorB	241
Concordance 6.4	Spokesperson of 'same-sex group': WenCorB	241
Concordance 6.5	Oppositional context of 'they': WenCorB	241

Abstract

This study aimed at analysing differences between the language use of men and women bloggers of Pakistani English e-newspapers. The corpus-based analysis compared two gendered corpora comprising a total of 11258 blog posts (6706 blog posts from 1674 men bloggers and 4552 blog posts from 1212 women bloggers) collected from blogs maintained by leading English e-newspapers of Pakistan. The data was collected from November 01, 2008 to August 31, 2015. Using 2015-version of the automated text analysis tool, Linguistic Inquiry and Word Count (LIWC), the study employed a quantitative top-down approach for obtaining results for 93 in-built language analysis categories of LIWC2015. The software processes each word in a text file, matches it to in-built dictionaries for 93 variables and groups results under several major categories with many sub-categories. The computed results were processed for statistical analysis with XLSTAT (Addinsoft, 2016) and MS Excel for normality test, two-tailed hypothesis test, Bonferroni's correction, and effect sizes. These tests were used to see statistically significant differences in gendered language use across 93 LIWC2015 variables. Besides, the study used AntConc 3.4.4 (Anthony, 2016), a stand-alone text analysis and concordancing toolkit, to note additional findings for differences in language use between men and women bloggers by comparing concordances, collocations and frequencies of selected expressions captured by LIWC2015 dictionaries. As a result, several gender differences in language use emerged. Many of these were consistent with the previous research conducted in other cultures and settings.

CHAPTER 1: INTRODUCTION

1.1 Chapter Overview

This chapter introduces the study in the following manner. First, it mentions the source of motivation for the present research, followed by background to the research study in hand. Then, the research problem is presented on which the research hypotheses are based. This is followed by rationale for the research. After that, the delimitation and significance of the study are presented. The chapter concludes by outlining the sequence of chapters in which the whole dissertation is divided.

1.2 Motivation for the Study

A potent motivating factor for the researcher to carry out the proposed study was a tutor-assessed course assignment, which he undertook while studying the module titled *Corpora in Applied Linguistics* as part of his PhD coursework in the Department of English, International Islamic University, Islamabad. The assignment prescribed an analysis of differences in language use of male and female bloggers. While many gendered blogs of native English varieties were available online, the researcher looked into exploring blogs representing Pakistani variety of English and Pakistani socio-cultural context. The blogs of English e-newspapers of Pakistan provided a viable option. Data from one such blog was gathered and analysed with the help of free online version of Linguistics Inquiry and Word Count (LIWC) tool available at http://liwc.wpengine.com, which provided interesting findings. Also, information on LIWC and relevant model for statistical analysis used in the current study were found in Newman, Groom, Handelman, and Pennebaker (2008) during search for relevant literature for the assignment. This

modest and small-scale success encouraged the researcher to analyse differences in gendered language at a larger scale in all available blogs of leading English e-newspapers of Pakistan.

1.3 Background to the Study

Language makes a significant distinction between human beings and animals and, within human languages, one sometimes hears the expressions "female language" and "male language" (Karlsson, 2007, p.4). Based on many differences found in the two genders and their language use, two prominent theories came to the fore: the *biological theory* and the *social constructionist theory*. Taking an *essentialist* approach, the former describes gender in terms of biological sex and assumes that men outsize and outpower women (Tannen, 1993; Bergvall, 1999); that gender polarities exist in language use; and, that gender roles are static and contextually independent. The latter is based on *non-essentialist* approach to gender differences and defines gender in the light of social contexts. It assumes that gender roles are fluid and contextually situated; that gendered identities are voluntary; and, that males and females choose their gendered identities in particular situations (Leaper & Smith, 2004). This theory further assumes that either gender is not confined to a particular language style, but exchanges styles based on the social context of interaction (Coates & Johnson, 2001).

1.4 Statement of Problem

The *social constructionist theory*—the main focus for many linguists nowadays— views gender as socially and culturally constructed and "something that is accomplished every time we speak" (Coates, 2004, p.7). This "fluid" approach (Speer, 2005, p.13) to language and gender challenges the *essentialist* view based on innate biological and psychological features that

differentiate men and women. This binary division of gender is at the root of the three pre-established approaches to language and gender: the *deficit approach,* with proponents like Otto Jespersen and Robin Lakoff; the *dominance approach*, supported by Dale Spender, Don Zimmerman, and Candace West; and the *difference approach*, attributed to Deborah Tannen. A preliminary look at the texts posted by the bloggers of English e-newspapers of Pakistan reveals that there is a significant difference in the language used by men and women bloggers, which calls for further examination to determine whether or not the language of these gendered blogs supports the *difference* approach. In this backdrop, this study is intended to analyse the written language of men and women bloggers of English e-newspapers in Pakistan.

1.5 Hypotheses

The study is based on the following hypotheses:

H_0: There is no difference in the language use of men and women bloggers of English e-newspapers of Pakistan.

H_1: There is a significant difference in the language use of men and women bloggers of English e-newspapers of Pakistan.

1.6 Justification of the Topic

Difference in the ways men and women use language has long been debated in sociolinguistics and discourse analysis. While some studies suggest that men and women use language differently, and, for different purposes, several studies have counter-argued the existence of any meaningful difference in men's and women's language (Newman, Groom, Handelman, & Pennebaker, 2008). However, much of the previous research in gender and language lacks

"specific context" (Li, 2014, 52), rely on small data sparse in naturally occurring language, which is drawn largely from limited population and apply qualitative analysis, which sometimes raise serious questions on the validity of their results—even if their findings conform to general expectations (Baker, 2014). Another significant aspect is that much of the previous research on language and gender has taken place in a foreign context, i.e. outside the context of Pakistan. Resultantly, there are still very limited studies on language and gender relationship in Pakistani context in general and in the domain of empirical research in particular.

One viable way to overcome the aforesaid shortcomings of previous research in language and gender, especially in context of Pakistan, is to use empirical research methods of corpus linguistics (CL) to analyse gender differences in language use. There is enormous potential in corpus-informed research to contribute to the field of language and gender by collecting a large data set of naturally occurring language from specific contexts and analysing it accurately through automated text analysis (ATA) tools, ensuring objectivity and avoiding bias in results. This study, therefore, applied corpus-based methods to analyse language differences between men and women bloggers of English e-newspapers of Pakistan.

This study will also contribute to the ongoing debate on gender-based language differences. The study considers the application of gender and language theories in the field, especially in the context of Pakistan. Since there is limited work available in Pakistan to explore gender-based differences in language use, with particular reference to the use of emerging *computer mediated communication* (CMC) and a boom in electronic media, wide-ranging efforts are needed to analyse new medium of personal expressions like blogs. The study explores this grossly under-researched phenomenon and provides insights from linguistic perspective. The study is likely to contribute to the following distinct areas:

(a) Testing the applicability of the claims of the *difference* approach to the texts of men and women bloggers of English e-newspapers in Pakistan.

(b) Providing valuable understanding of the gender-based differences in the language used in Pakistani English e-newspaper blogs.

(c) Adding to research methodology in Pakistan in the field of language and gender with reference to gender-based studies in sociolinguistics and CL.

1.7 Delimitation of the Study

To analyse differences in gendered language, this study has been limited to blogs of English e-newspapers in Pakistan because blogs emerge to be a particular kind of web-genre where writers feel free to express and share their ideas, feelings and thoughts using a variety of language resources. Two separate specialised corpora have been built for the study: one for men bloggers (called Men Corpus of Bloggers or MenCorB); the other for women bloggers (called Women Corpus of Bloggers or WenCorB). Both the corpora (termed together as Gendered Corpora of Bloggers or GenCorB) were built by retrieving blog posts from five leading English e-newspapers of Pakistan, namely (in alphabetic order) *Dawn* (D), *The Daily Times* (TDT), *The Express Tribune* (TExT), *The Nation* (TNn), and *The News* (TNw).

In all, 11258 posts were included in this study. Of these, 6706 posts were of 1674 men bloggers whereas the remaining 4552 posts were of 1212 women bloggers. The study included blogs posted between November 01, 2008 and August 31, 2015. Based on a criterion for selection of gendered blog posts for this study elaborated below, the posts were retrieved from online

archives of all the five Pakistani English e-newspapers. Following a chronological sequence here, online archives of D provided access to blog posts included in this study from November 01, 2008; TExT from January 01, 2010; TNw from September 24, 2011; TNn from July 03, 2013; and, TDT from April 01, 2015.

Blog posts only by adult Pakistani bloggers were included in the study while those by foreign nationals and minors, i.e. below 18 years of age, were excluded, as in Pakistan the age of adulthood is 18 years (Government of Pakistan, 2006). Bloggers' profiles available with each blog post were used for determining the nationality and age of bloggers. Blog posts, which carried doubtful or vague gender identity, e.g. posts by bloggers who used the word 'anonymous', pennames, names of organisations or blog-group names as their identity, were excluded. Similarly, blog posts jointly contributed by a man and a woman were also not included for reason of vague authorship identity.

The retrieval and storage of GenCorB comprising about 9.06 million words (MencCoB = 5.6 and WenCorB 3.3 million words respectively) single-handedly was an enormous undertaking. Thanks to the web archives of all the e-newspapers, the blog posts were readily available online. In almost all cases, weblinks were working perfectly well. Yet, the entire data was collected in one and a half years, given the large quantity of data and the time available with the researcher for data collection. For data analysis, the study employed two ATA tools, viz, LIWC2015 and AntConc 3.4.4 (Anthony, 2016) alongside XLSTAT (Addinsoft, 2016) and MS Excel for statistical analyses. Complete sampling criterion and data analysis procedures have been explained in *Chapter 3: Data & Methods* of this study.

1.8 Significance of the Study

This study is significant for several reasons. Firstly, there is still a lot of scope for research to be carried out in the context of Pakistan in the proposed field. Although studies have been carried out in the area of language and gender, much of those have taken place in linguistic, social and cultural contexts outside Pakistan, focussing on the differences of language use by the native-speakers' varieties of English. There is a recent shift to studying non-native varieties of English in research on linguistics in general (Rahman, 2010) and analysing more "localised studies" in language and gender research in particular (Swann, 2002, p. 59). As a result, future studies would be required to carry out small-scale localised language analysis and relate their results to wider social contexts (Gormley, 2015). The proposed research will, therefore, be a useful addition to studying gender-based differences in language use in the linguistic, social and cultural context of Pakistan.

Secondly, in today's digital, fast-progressing and technology-intensive world, CMC has emerged as a new form of language (Baker, 2010a). As part of this digital boom, digital journalism (i.e. communicating along with discussing news and events round the globe through social networking websites, e-newspapers and blogging) is gradually becoming the need of the day (Kaukab & Mehrunnisa, 2014). E-newspaper blogs are increasingly used for effective communication of views on a vast array of topics by both male and female bloggers. Some previous studies (e.g. Rodino, 1997), disagree with the conceptualisation of gendered language use in terms of binary opposition and argues that traditional gender and language research oversimplifies language online because being *virtual* allows more freedom and flexibility. It is, therefore, important to see if new CMC contexts resonate with the linguistic findings of language

use by the genders given in the past studies or reveal new ones. The present study would be a useful endeavour in this direction.

Thirdly, women have been playing a significant role in Pakistan in every walk of life. Effective communication between couples in a house as well as between members of male/female genders as employers/employees or bosses/subordinates in all professions is *sine qua non* for success. In all these varied settings, if male language is taken as a norm, female language can be viewed as totally different. Lack of understanding in male and female language use can resultantly pose serious hurdles in effective communication between members of the two genders. This research will be a useful addition to the existing body of knowledge that focuses on better understanding of these differences to improve effectiveness of communication between genders in different settings.

Fourthly, this research may have commercial value in the sense that user(s) of this study can find *which* topics, services, and/or products are discussed, used and liked/disliked by male and female in Pakistan, as represented by the e-newspaper bloggers. The findings of the study may, therefore, be helpful in consumer profiling in the country.

Fifthly, the study also has significance for future research in that it will highlight gender-based language features, which can help researchers analyse written texts for gender-appropriate language in both fiction and non-fiction texts. Fictional characters represent language use by particular groups (Baker, 2010b). Gender-inappropriate words put in the mouth of characters seldom appeal to the readers of fiction or viewers of plays. For instance, quoting a study by Krammer (1974), who analysed women representation in cartoons of *The New Yorker* magazine published between February 17 and May 12, 1973, Wardhaugh (2006) remarked that though cartoons do not record actual speech, they must be representative of the speech, which,

people believe, occurs in real-life. Similarly, for analysis of non-fiction context, this research would be useful as historically significant but anonymous texts can be examined in the light of gendered linguistic features reported by such research to determine an unknown author's gender. Another important area where this study may contribute is forensic linguistics, especially in determining the gender of suspected authors, which can assist court trials and criminal investigations. In this regard, Rustagi, Prasath, Goswani, & Sarkar (2009), Prasath (2010), and Fatima, Hasan, Anwar, & Nawab (2017) are relevant instances that show how language features in texts can be used for gender identification in the filed of forensic linguistics.

Finally, this study may highlight the socio-cultural dynamics of the Pakistani society. All sociolinguistics research is primarily concerned with establishing a link between the social context and the language use (Bloome & Greene, 2002). E-newspaper blogs are meant for public knowledge and general readership, which, on the one hand, represent the social context in which the texts have been created and, on the other hand, influence that social context by potentially influencing the language used in the society. The study of e-newspaper blogs, therefore, has the potential to reveal the social and cultural patterns of thinking and living of Pakistani male and female bloggers as well as the social and cultural challenges, which they captured in their writings on a variety of topics.

1.9 Outline of Chapters

This thesis has been arranged in the following manner: The first chapter (*Introduction*) discusses what motivated the researcher to undertake this study, followed by problematising and contextualising the study, framing specific research hypotheses, justifying the topic, explaining the study's significance, stating the study objectives, and delimiting the study. The second chapter

(*Literature Review*) presents related literature and discusses different dimensions of the research topic—assigning special focus to language and gender research—followed by a short history of newspapers in Pakistan and the advent of e-newspapers. Subsequently, the chapter discusses blogging, e-newspaper blogging and different methodologies for analysing newspaper contents. The chapter concludes with a review of language and gender research in Pakistan. The third chapter (*Data & Methods*) explains the research methodology adopted for the study. It discusses, at length, and, in a step-by-step manner, the process of data collection, data organisation and data analysis. The fourth chapter (*Results*) explains, in a systematic way, how various measures were applied to analysis of the data, presents the results for this study and determines whether or not the results support the hypotheses. The fifth chapter (*Discussions*), alongside indicating any additional findings, compares the results of the present study with the findings by previous research. The results are presented with the help of tables and figures. This is followed by the interpretation of the results. The final chapter (*Conclusions*) summarises the findings and limitations of the study and suggests scope for further research in the area.

CHAPTER 2: LITERATURE REVIEW

2.1 Chapter Overview

This chapter presents an overview of literature relevant to the current study. Since the topic of the study involves an overlap of literature in several areas, this review has been divided, for readers' convenience, into four main sections—*theory*, *method*, *genre* and *context*—with sub-sections. The *theory* section traces a brief historical background of language and gender studies, leads into discussion on four theoretical frameworks for studying language and gender and concludes with a rationale behind choosing the *difference* approach for the present study. The *method* section reviews CL studies and tools that have been used to explore language and gender relationship, mainly the differences in gendered linguistics features. This section is followed by an introductory discussion on the *genre*, i.e. weblogs, and studies on gendered language differences in weblogs. The next section highlights the country *context* in terms of sex-ratio in demography, language and gender research as well as Pakistani blogosphere. While gaps for the present study have been implicitly indicated in each of these sections, a brief but explicit note has been added as the final section of this review.

2.2 Theory: Language and Gender

Though differences in the language of men and women is reported in publications as early as Vives' *De Institutione Christianae Feminae* (On the Instruction of a Christian Woman) published in 1523 during the Tudor period (1485-1603), the systematic discipline developed much later in 1960s (Kolsky, 2012). A detailed overview of such early instances of studies on language and gender—especially the language of women—has been presented by Grey (2010). However, a more systematic approach to studying language and gender appeared in the 1960s and 1970s in

the studies of a renowned linguist, William Labov. His work is considered as a pioneering attempt in sociolinguistics in the field of language variation.

Labov (1966, 1972a, 1972b) worked on social stratification of linguistic variables, especially the social class and geographic location of the members of the population in which he compared the standard (prestigious) and nonstandard forms of speakers. What more significant came to the fore from his studies was that, alongside class, gender was also considered as influencing speakers' phonological and grammatical features. His studies highlighted that, in each social class, women used more prestigious forms than men and that this feature was significantly telling in the lower middle-class grouping. Identifying this tendency as hypercorrection or overuse of prestige forms, Labov argued that women of lower middle class used more prestigious form because they felt insecure about their status. For men, he found, there appeared to be relation of masculinity with nonstandard forms of language in male working class groups, which suggested that covert prestige was associated with vernacular forms.

The notion of overt and covert prestige was further investigated by Trudgill (1972) whose findings supported Labov's conclusions. He also found in his study in Norwich (UK) that male speakers used nonstandard forms more frequently in comparison to female speakers. He argued, like Labov, that this difference in language use between the two genders is due to status consciousness: women are more status-conscious and they feel socially less secure than men.

Quite alternative interpretations of and explanation for gender differences in language use were given by the *social network approach* adopted by Milroy and Milroy (1978), and Milroy (1980). They found that the difference in use of prestigious/vernacular forms of language by members of both the genders was a result of their group membership. The approach, therefore,

suggested that instead of social class, it is the group pressure that compels individual members to follow language norms.

Until this timeframe, however, gender was regarded as a sociolinguistic variable, just like social class, age, ethnicity and social status. It was not until the mid-1970s when Lakoff's (1975) seminal essay *Language and Women's Place* was published and the science and the field of gender and language took a new shape (Karlsson, 2007).

2.2.1 The Field Takes a New Shape

For centuries, the relationship between language and gender was understood before this relationship was considered worthy of scientific study. This understanding was based on both prescriptive ideas, i.e. "how women and men should speak", and "folk-linguistic" ideas about how they actually do (Sunderland, 2008, p.2). The folk-linguistic ideas were captured in the proverbs, which often than not, referred disparagingly to women's verbosity. However, in such kinds of early descriptions of "women language", women's verbosity was not compared to that of men but to silence. As a result, any talk by women was regarded as too much (Spender, 1980).

The year 1975 was a watershed in launching the field of language and gender, which witnessed the publication of three influential books: Robin Lakoff's *Language and Woman's Place*, Mary Ritchie Key's *Male/Female Language*, and Barrie Thorne and Nancy Henley's edited volume *Language and Sex: Difference and Dominance*. These ground-breaking works appeared during the 1970s' feminist movement when scholars had begun to challenge male-dominated norms, and the *essentialist* view of difference in behaviour of women and men as an outcome of their biological differences (Kendall & Tannen, 2001). Also, a conceptual divide was theorized between biological sex and socio-cultural constructs of gender.

The following decade saw scholars refining and advancing understanding about gender and discourse relationship. Studies focussed on talk among women, narrative, language socialization, language among children and adolescents, and language and gender in particular contexts, e.g. doctor–patient interaction. After 1990s, the field grew with the publication of numerous edited collections. As the understanding of the relationship between language and gender progressed, researchers arrived at many similar conclusions, although these similarities frequently went unrecognized or unacknowledged. The most widely debated issues, however, remained *gender duality*, i.e. the division of speech on the basis of a binary division of gender or sex and *performativity*, i.e. local practices bring gender into being (Kendall & Tannen, 2001).

2.2.2 Theorising Language and Gender

Otto Jespersen—a professor in the University of Copenhagen—was the first to publish a piece in modern linguistics as regards "women's language" (Githens, 1996). In his book *Language: Its Nature, Development and Origin*, Jespersen (1922) devoted Chapter XIII entitled "The Woman", to describing differences in women's language compared to men's where he noted that the difference between men's and women's language is viewed in many countries of the world as a peaceful struggle for dominance. He analysed differences in men's and women's language in conversation, phonetics and grammar, choice of words, vocabulary, adverbs, periods and, finally, related women's language to her thoughts while commenting upon "general characteristics" of women's language (Jespersen, 1922, p. 252). He attributed language differences between men and women to the early division of labour between the two genders. While concluding, he expressed the hope that the social changes taking place at the time may finally bring about a change in the

linguistic relationship between the two genders. Quite conceivably, he predicted the very language differences that linguists today are looking for in even new settings and contexts.

Following early works like Jespersen's (1922), studies, debates and critiques by prominent linguists in the field of language and gender, such as Robin Lakoff, Deborah Tannen and Deborah Cameron gave birth to prominent theoretical approaches to the study of relationship between language and gender. Taking the risk of oversimplification, four theoretical approaches can be identified, namely the *deficit*, *dominance*, *difference* and *dynamic* approaches. Among these, the first three approaches, also called as *3Ds*, were rooted in the *essentialist* view based on *Biological Theory* of gender polarity whereas the fourth approach has its roots in the *Social Constructionist Theory* of gender fluidity. Although these approaches developed in a chronological and historical sequence, it does not mean that one approach superseded the other; rather, they remained at any given time together, though in hot contention (Coates, 2007).

The *Deficit* Approach

The founding proponent of the *deficit* approach is Robin Lakoff who described male language to be more prestigious, more desirable, and stronger in comparison to female language. She argues that women socialise in a more ladylike manner, which bears direct impact on their linguistic behaviour too; therefore, they remain to be less powerful in culture. Earlier, she had indicated women's deficiency in language behaviour as they are constantly scorned for their 'inferior' language ability in comparison to men (Lakoff, 1973).

As the first theoretical standpoint launched in early 1970s, this approach views women as disadvantaged in language use, with their language not conforming to the dominant but implicit male norm of speech. Lakoff (1973) argues that with uncertain expression, extreme deference

and politeness, women make their language different from and inferior to men's language. Lakoff (1975) includes ten assumptions about what she felt makes women's language: *hedges, (super)polite forms, tag questions, speaking in italics, empty adjectives, hypercorrect grammar and pronunciation, sense of humour lacking, direct quotation, special lexicon* and *question intonation* in declarative statements. Much of her observations conformed to those of Jespersen's (1922).

This approach, however, was challenged on the grounds that it looks at women's language as there is something "intrinsically wrong" (Coates, 2007, p. 56) with it and, therefore, it must conform to male language norms to be correct. For some time, this approach was used in "assertiveness training" for women (Talbot, 2010, p. 98), which is now viewed as "outdated" in research studies (Coates, 2007, p. 66).

The *Dominance* Approach

Lakoff (1973) presented her work as a starting point for future research. Her work, therefore, prompted discussion on another approach to the study of men's and women's language, i.e. the *dominance* approach, which is associated with Zimmerman and West (1975), Fishman (1978), and Spender (1980). This approach views women as the oppressed gender and explains gendered language differences in terms of men's dominant role and women's subordinate role (Fishman, 1997). Studies following this approach focused on how linguistic pattern determines male dominance. In the light of this approach, "doing power" became synonymous to "doing gender" (West & Zimmermann, 1983). The approach further argues that all members of society—men and women—contribute to maintaining male dominance in female oppression.

Famous works in this approach are Fishman's (1978) paper titled *Interaction: The Work Women Do* in which she published her observations of interactions between three couples at home. In this study, she focused on the conversational division of labour of the genders. She conducted her study in a domestic environment and observed that wives worked hard and did a lot of support work to continue conversation with their husbands. She generalised her observations to a much larger social order in everyday interactions between the genders that prevail in the society and concluded that since there is an unequal division of work-sharing in the society on the basis of gender, the labour in private conversation is also unequally divided. Another famous work, well-grounded in this approach, is Spender's (1980) *Man Made Language* in which she presents a monolithic view of male power, i.e. as if all men are in a position to dominate all women, which may actually not be the case.

The *Difference* Approach

The *difference* approach is often offered as an alternative to the *dominance* approach to explain language differences between men and women. Therefore, behaviour that was previously interpreted as men's desire to dominate women is now seen as the outcome of their upbringing in different gender-specific sub-cultures (Coates, 2007).

The initial study of language patterns used by men and women under the *difference* approach is attributed to John Gumperz—a professor of anthropology in the University of California, Berkely—and his colleagues (Talbot, 2010). In *Crosstalk*—a sub-programme of a series of ten programmes titled *Multi-Racial Britain*—which was first broadcast from British Broadcasting Corporation (BBC) television on May 1, 1979, he introduced his pioneering work in the field of applied sociolinguistics. His focus in *Crosstalk* was the issue of workplace

miscommunication on the basis of racial and ethnic stratification. His latter insights on gender inequality and other research contributions stemmed from this programme (The Regents of the University of California, 2007). In this talk show, he mentioned his observations of language differences in boys and girls in the context of cross-cultural miscommunication. His model was based on two distinct gender-specific cultures for boys and girls in which they grow.

Later, Maltz and Broker (1982) developed their study on cross-cultural miscommunication on the foundations set by Gumperz and Goodwin's observations of linguistic dimensions of play among children from black community in Philadelphia in Goodwin's three famous works (Goodwin, 1978, 1980a, 1980b). Maltz and Broker (1982) based their argument on the grounds that since boys and girls grow up in different gender-specific sub-cultures that have distinct rules for speaking, their acquisition of different norms of engaging and interpreting conversation is inevitable (as cited in Aries, 1996).

It was Deborah Tannen—a linguistics professor at Georgetown University—who built on this argument. Being Lakoff's student, she was introduced to her work on language and gender. Tannen (1986) had already published a book on conversational styles titled *That's Not What I Meant! How Conversational Style Makes or Breaks Your Relations with Others* in which she had discussed conversational styles in domestic life at home and how differences in conversational styles were marring close relationships at home.

After receiving much popular response to this early effort, Tannen (1990b) decided to write in detail on gender and language differences, hence her publication *You Just Don't Understand: Women and Men in Conversation*. She argued that there are contrasting language styles in genders that men and women need to understand to avoid blaming each other or their relationship. She looked at these language differences in social perspective as she believed that since the upbringing

of boys and girls takes place in separate 'cultures', their talk is essentially a manifestation of "cross-cultural communication" (1990b, 18).

She gave another important distinction between language of men and women, i.e. "rapport talk" and "report talk". She claimed that women use language for intimacy (rapport-talk) while men use language for information (report-talk). Tannen (1990b) also contributed six major areas of contrasts to represent language use differences between men and women, which she has explained with the help of relevant examples. These are *status vs support* (i.e. men's communication is competitive while women communicate to seek consensus and support); *independence vs intimacy* (i.e. women think of support and struggle to preserve intimacy while men, concerned with status, want independence); *advice vs understanding* (i.e. while men look to problems as challenges and look for their solutions, women seek emotional support and sympathy); *information vs feelings* (i.e. men consider communication as a source of conveying information whereas women attach feelings to communication); *order vs proposal* (i.e. while men communicate in a more direct way/imperatives, women communicate through suggestions using expressions like "let's"); and, *conflict vs compromise* (i.e. while men habitually oppose others and can adjust their style to the nature of conflicts, women generally avoid opposing the will of others).

Since Tannen (2010) was much concerned with developing an understanding between couples, she concluded that men and women need to learn each other's language. She clearly distanced herself from the *dominance* approach and held that given the stylistic difference between the two genders, a miscommunication was inevitable and the key to avoiding this miscommunication is developing an understanding.

As Cameron (as cited in Talbot, 2010, p. 99) observed, *dominance* and *difference* approaches represented feminist movements: the former was "the movement of women's outrage"

while the latter "was the movement of feminist celebration. . . the revaluing of women's distinctive cultural traditions". The *difference* model is currently widely used in the world, in various institutions besides academia. It also has popularity with journalists and mass media writers. In addition, it has been used in Britain by telephone companies to stimulate social use of the phone while some British advertising companies used it to show contrast in men's and women's use of phone, showing women as good communicators (Talbot, 2010).

The *Dynamic* Approach

The fourth and most recently developed approach to language and gender is the *dynamic* approach, also known as the *social constructionist* approach. This approach views gender identity as something socially constructed rather than a predetermined construct (Coates, 2007). In this approach, as West and Zimmerman (1987) argued, speakers are seen as "doing gender" rather than "being" a particular gender. The "doing gender" expression was also used by the *dominance* approach; the reason, however, is that none of the four approaches has developed in isolation. Rather, one is influenced by the other; so, there is a visible overlap among the four approaches.

The *dynamic* approach does not view gender as a static or essential characteristic of a speaker but something that speakers accomplish every time they talk. Therefore, this approach is in sharp contrast to the *essentialist* view of gender adopted by the early three models: *deficit*, *dominance*, and *difference*. The dynamic approach shifts gender from a noun to a verb and views it not from the perspective of something learnt in early childhood but something that is constantly performed. Within the broader framework of the *dynamic* approach, four sub-approaches merit discussion: *performativity*, *community of practice* (CoP), *indirect indexing* and *gendered discourse*.

Butler's (1990, 1993) work on gender and *performativity* has been particularly significant in recent research who thinks that gender is a repeated performance. She, therefore, holds that gender is not a possession but a process, which an individual has to constantly perform. However, stress on performativity for her does not mean that individuals can be anything at any time they want to. In her views, the category of gender already exists before individuals perform it; however, individuals only negotiate with gender when they perform their identities. According to her, gender is a performative social construct or a repeated sequence of performances over a range of behaviour associated with a particular sex, and, therefore, a process that constantly takes place in line with pre-existing gender norms that individuals negotiate.

Performativity approach is a clear departure from the view of a binary opposition of genders to something more conceptually sophisticated interpretation of gender, which has led to more complex nuances and concerns with the diverse ways individuals negotiate with, contest and reaffirm what they think as appropriate representation of gender in their behaviour in particular contexts.

For developing gender identity in particular contexts, the *CoP* approach is of special interest for linguists. The first proponents of this approach in the field of language and gender were Eckert and McConnell-Ginet (1992a, 1992b). They borrow Jean Lave and Etienne Wenger's notion of "community of practice" by which Eckert and McConnell-Ginet mean a community based on "social engagement", instead of particular locality or population, and the language that serves such a social engagement (1992b, p. 94). From the perspective of *CoP*, individuals belong to a variety of 'communities': at work, at home, at hobbies etc. Therefore, a person's gender identity is dynamic, fluid and variable, which produced and reproduced during social interaction within these communities.

In interactional studies on language and gender, the notion of *indirect indexicality* has been recently popular. Ochs (1992), the chief proponent of this approach, argues that the relationship between language and gender is established through a web of meaning in society. She differentiates between two types of indices: *direct indexicality,* which encodes gender explicitly like Mr/Mrs, man/women, and *indirect indexicality,* where particular language styles determine genders, e.g. association of women with tag questions and politeness or men with coarseness.

In the domain of *gendered discourse*, the poststructuralist theory has been more influential, particularly the works of Foucault (1972, 1979, 1981) and his beliefs on discourse as a social practice. Foucault (1979) believes that discourse structures our sense of reality—of self, of objects, of world, of others—and human beings impose these structures on objects and experiences when interpreting them. With regard to gender, Mills (2004) argues that discourses pertaining to masculinity and femininity delineate the limits within which men and women perform their gender identities. Basing their contention on Foucault's line of argument, Cameron and Kulick (2003, p. 16) challenge the "language-dependent practices of definition, classification, explanation, and justification"; or, alternatively, bodies of knowledge that are socially constructed.

Together the different models within the broader theoretical framework of the *dynamic* approach believe that there exists a potential multiplicity of masculinities and femininities that individuals negotiate, accept or reject, which has created a new space for further research to conduct more sophisticated analyses of how gender is constructed in both spoken and written data.

The 1990s saw paradigm shifts in academic interpretation of gender, especially with the development of the *dynamic* approach, which triggered further debates on the relevancy of the other approaches to language and gender. The basic contention between the *essentialist* approaches (*deficit*, *dominance*, *difference*) and the non-essentialist approach (*dynamic*) is that the former

considers that language is the product of gender while the latter thinks that gender is the product of language.

While the *deficit* approach is now considered as dated, the remaining three approaches, i.e. *dominance*, *difference* and *dynamic*, have contributed significant insights into the nature of gender differences in language (Coates, 2007) and are, therefore, still employed in research.

Theoretical Underpinning: Present Research

One strong motivation for current research comes from the fact that research in the field of language and gender has largely concentrated on English-speaking societies. More recently, researchers feel encouraged to study language patterns of men and women in a variety of societies and cultures. There is now an overriding emphasis on the fact that gender is to be viewed more locally and that it interacts more closely with the social environment in which it is studied. This has enabled researchers to widen the scope of studies not only to non-English speaking societies but also to particular contexts of gender interactions to study the relevance of different theoretical frameworks to these settings. The present study is also an attempt in the same direction to see the relevance of the *difference* approach to language and gender within the cultural environment of Pakistan. This approach has been considered appropriate for the present study for several reasons.

First, as highlighted above, there is a need to widen the scope of research based on various approaches to language and gender to newer linguistic and social contexts. From linguistic point of view, Pakistani context is unique. Pakistani English has evolved as a special variety of English, which corresponds to Standard British English (SBE) in relation to linguistic patterns (Khan, 2012). In this backdrop, it is deemed appropriate to extend the insights of the *difference* approach in language and gender for analysis of the Pakistani variety of English.

Second, in language and gender studies, gender is not something too simplistic. It has close interaction with age, class, social status and culture, in addition to the spoken and written genres of gendered language. Therefore, findings of studies conducted in a particular context may not be generalisable to newer contexts of gendered language. Hence, there is a need to carry out studies in different specific contexts to create a large photomosaic of relationship between language and gender.

Third, Pakistani cultural context has close similarity to Tannen's (1990b) notion of two-cultures. In Pakistan, children predominantly grow in gender-specific cultures. Following the Eastern outlook of integrated family life, men and women, who come from different families, do not integrate much in public spheres in Pakistan unlike many modern Western societies. The institutions of family, clan, ethnicity, religion, and class also have a major part in determining the social behaviour of the members of Pakistani society. With the passage of time, the structures of these institutions are becoming 'modern', but such modernisation takes place within the grooves of already established traditions.

Finally, unlike the assumptions of *deficit* and *dominance* approaches, the *difference* approach is based on 'equal-but-different' notion of genders. In Pakistani society, in the light of the Article 25(2) of the Constitution of Pakistan—the country's supreme law—legally women enjoy an equal status to men (Government of Pakistan, 1973). Although apparently patriarchy may seem to dominate the lives of women in Pakistan, a closer examination of the Pakistani family life—the basic unit of society—is likely to reveal a more different but subtle reality. As pointed out by some research (Qadeer, 2006), within the patriarchal framework of families in Pakistan, women carve out a personal space for them in two ways: First, women become more authoritative in families as dominating mothers and mothers-in-law in later years of life from dominated

daughters and daughters-in-law in their early age. Moreover, gender relations often reverse: men become more dependent on women for care and support in later years of their life. Second, women of middle- and upper-middle classes and those living in cities enjoy more freedom compared to members of the same gender from poor or backward social settings. The prestige of high-class family status rubs off on women elevating their social status and giving them more liberty and space for gender equality.

In the light of the above discussion, in Pakistan's socio-cultural context, the *difference* approach seems to be suitable for the present study: it is neither based on superior-inferior biases between genders advocated by the *deficit* and *dominance* approaches nor does it make gender a matter of choice for individuals (as done by *dynamic* approach). The *difference* approach is distinct in that it suggests that language difference in genders is simply the outcome of internalization of different norms of communication.

2.3 Method: Corpus Linguistics

In contrast to the traditional linguistic approaches to studying language with rather limited and intuitive language data conceived and produced by the linguist(s), CL approaches study language by analysing frequency and language patterns in a comparatively larger body of naturally occurring language texts. The emergence of computers has revolutionised CL research since 1960s. An instance of this is Brown University Standardised Corpus of Present-Day English (or the Brown Corpus), a collection of one million words of American English from 15 different text types published in 1961, which is regarded as the starting point of CL.

CL is both a branch of linguistics and methodology, or, in other words, a science and a methodology. Like a science or a field of linguistics, it uses observations, experimentation

and measurement; it follows its own approaches to study language. As a methodology, it has its own set procedures of data collection, analysis and interpretation (McEnery & Wilson, 1996; Baker, 2011).

CL studies have several merits: First, a large body of language can be analysed at once. For instance, even before the advent of computers and modern CL techniques, Käding (1897, as cited in McEnery & Wilson, 1996) analysed manually a German corpus of about 11 million words for spelling conventions. Secondly, the patterns in linguistic data or *corpus* are subjected to measurement techniques grounded in statistics, which can be verified through replication. Finally, the use of computers in modern times for CL research makes sorting, retrieving and calculation of linguistic data easier and more accurate.

While CL research has these advantages, it also has limitations as pointed out by Bennett (2010). First, it can tell language instances that are present in the corpus, but it cannot show what is outside it, i.e. the language that is not used in a corpus. Second, it can only show *what is* but cannot explain it. The linguist has to interpret it. Lastly, while it represents a large body of language, no corpus can claim to have encompassed all the language, i.e. the language that a corpus contains is not random but planned.

According to Baker (2010b), CL research has contributed to three major aspects of language and gender research. The first of these is the analysis of gender differences or similarities in language use. In this context, prominent instances include studies by Rayson, Leech, and Hodges (1997), who compared gender differences in lexical choice using the British National Corpus (BNC), and Harrington (2008), who compared gendered language on the basis of use of reported speech. The second aspect is addition to the knowledge about gender representation in society. In this respect, for instance, studies analyse how men and women are talked or written

about (e.g. Kjellmer, 1986; Biber et al, 1999). The third aspect is the male bias in language use, i.e. to see how frequently men and women have been represented in a text (e.g. Baker, 2010b). The subsequent part of this review is dedicated to the first aspect highlighted by Baker (2010b) for its relevance to the present study.

2.3.1 CL Studies: Difference in Language Use by Genders

A significant contribution to the study of language and gender relationship using CL techniques, though conducted on a small scale, was Biber, Conrad, and Reppen (1998). The study examined a corpus of 276 personal letters written by men and women, which were categorised across countries and further grouped on the basis of male/female recipients (M-M; F-M; F-F; and M-F) to verify the claims of difference in style of genders. It tested two claims: First, the use of more emphatic forms by genders and concluded, without any statistical testing, that women used more emphatic forms than men, especially in the 20th century. The second claim to be tested was to see whether Dimension 1 variable separates gender groups. Dimension 1 is the result of factor analysis. According to Biber (1988), it can distinguish between different registers of writing and speech. The variable consists of two poles: *involved* pole, which includes pronouns, mental verbs, sentence relatives and demonstratives; and, *informational* pole represented by language features such as prepositions, word length, adjectives and nouns. When Biber, Conrad, and Reppen (1998) applied Dimension 1 to the corpus of letters, they found men more *involved* and women more *informational.* They also observed an accommodation effect and reported that men were more *involved* in letters addressed to women in comparison to their letters addressed to their own sex group.

The application of Dimension 1 to language and gender difference was also tested by other studies. For instance, to verify previous findings of sociolinguistics research, especially Tannen's *difference* model, Heylighen and Dewaele (2002), built a corpus of about 30,000 words of two speaking styles and one writing style in decreasing order of contextuality: informal conversation, oral examination and examination essays. They researchers suggested a new way of measuring *deep formality*, which closely corresponded to Bibers's (1988) Dimension 1. They defined *deep formality* as the "attention to form for the sake of unequivocal understanding of the precise meaning of the expression" (Heylighen & Dewaele, 1999, p.3). After reviewing literature on the concept of *deep formality* as a pervasive and main linguistic feature of language variation across languages, they assessed its correlation with gender. In this empirical study, they concluded women were less formal than men and claimed that this difference applied to a large number of languages. They, however, found that genre also played an important role in language formality as the difference between genders disappeared in formal writings like essays. They explained this phenomenon by citing psychological and sociolinguistic evidence of women being generally more *involved* in conversation as compared to men who are more *informational.* They also presented cognitive evidence in this explanation, which supports the hypothesis that men and women have different cognitive orientation.

A related study to Dimension 1 or *deep formality* was conducted by Schmid (2003), who studied spoken text in the BNC to verify whether the two genders live in different cultures as hypothesised by Lakoff (1975). He calculated frequency for a number of words that previous research found gendered. His results were consistent with those of the other researchers: females used more colour words, hedges, and temporal adverbs whereas few abstract nouns in comparison to men. He also concluded that women were more engaged and *involved* in spontaneous speech.

The discussion on linguistic features of language in the *involved-informational* dichotomy and its correlation to gender was again discovered by Saily, Siirtola, and Nevalainen (2011). Their study analysed part-of-speech (POS) tagged Corpus of Early English Correspondence, which consisted of more than two million words produced by 660 men and women in their personal letters. Primarily developed and designed for diachronic linguistic studies, the corpus, nevertheless, rendered itself suitably well for a sociolinguistic investigation of gender and genre on frequency count of nouns and pronouns.

In this study, the researchers concluded that men used more *informational* (nominal) style whereas women's style tends to be more *involved* (pronominal). Unlike the study of Biber, Conrad, and Reppen (1998), this study, used statistical techniques and found that, barring for one year, the difference was significant over a wide span of time between the years 1415 A.D. and 1681 A.D. Nevertheless, the researchers admitted that the study did not control variables like the writer's level of education, their social class, the letters' topics and the sender-receiver relationship, which might have influenced the results together with other social parameters.

In CL research on differences in language use by men and women, some studies also investigated the areas of swear words. Once such analysis was conducted by McEnery (2006) who examined the relationship between the speakers' gender, the words used in swearing and the listener's gender. He recorded some significant observations as an outcome of this study, which included that use of swearing words was more frequent in the same gender dyads as compared to dyads from different genders. Similarly, he observed that there was preference in choice of certain swearing words for genders. In general, it appeared that women were less likely to hear swearing words than men, especially strong swearing words.

The tendency of swearing words was also studied by Thelwall (2008) among UK and US women and men of young and middle ages. He studied a social networking site *MySpcace*, which is you-oriented and was found to be most visited by the US youngsters besides enjoying a wider user-base in the UK. He found that while there was strong difference in the use of swearing words by genders in the US—males used more swearing words than females—there tended to be no difference in the use of swearing words by genders in UK. He attributed this to the rise of the *ladette* culture in UK, an expression used for "a young woman who behaves in a boisterously assertive or crude manner and engages in heavy drinking sessions" (ladette, 2017).

The observations of McEnery's (2006) were, nevertheless, contrary to Rayson, Leech, and Hodges (1997). He found that swearing words were less common in women's speech in their sub-corpus from BNC containing 10 million words. His study also confirmed some of the findings of previous studies by observing that men's speech was more *informational* in that it used more determiners, the preposition *of* and numbers whereas women's speech was relatively more *involved* as it was more pronominal. In addition, he found that women use more proper nouns overall, especially those referring to people, while men used proper nouns referring to places.

Within language and gender research based on BNC, some studies (Koppel, Argamon, & Shimoni, 2002; Argamon, Koppel, & Fine, 2003) have also examined difference in use of language by genders in fiction and non-fiction sub-corpora. What distinguished these studies was an important gap, which the authors identified in the previous gender language studies as being limited to analysis of linguistic features in speech genre only with a very few focussing on the informal written genres. They argued that speech allows for cues—e.g. intonational cues—which do not appear in the written text. Even studies using transcribed speech (Rayson, Leech, & Hodges, 1997) could not ignore certain spoken language features such as *fillers* (e.g. *err* or *umm*). These

studies concluded that while prepositions were generally used more by men, women tended to use two prepositions, i.e. *for* and *with*, more significantly. In the same way, men used more *he*, whereas women scored higher on the use of overall pronouns. These results again corresponded to Biber's (1988) *involved-informational* dichotomy.

2.3.2 Corpus Tools

There may be other tools and techniques used in CL to analyse language use by men and women, this study, however, applied two of them, LIWC2015 and AntConc 3.4.4 (Anthony, 2016). The former was used for a comparison of differences in language use by men and women bloggers across its 93 in-built categories; the latter complemented the LIWC2015 results by analysis of differences in frequencies, concordances and collocations of particular expressions captured by LIWC2015 in GenCorB. This section of the review, therefore, covers the application of LIWC by previous research to the study of differences in language use by men and women.

LIWC's Development

LIWC (pronounced *Luke)* was initially developed in 1993. It was an outcome of an exploratory study conducted in the same year by James W. Pennebaker, a professor and social psychologist at the University of Texas at Austin, and his associate, Martha E. Francis. Ever since, the programme has been regularly updated with addition of some new features in 2001, 2007 and 2015. In the second (LIWC2001) and third versions (LIWC2007), the original application was improved with an expansion in its dictionaries. The latest version of the package, also used for the present study (i.e. LIWC2015) has significant upgrades in dictionary and software options and offers 93 output variables to categorise various language features (Pennebaker, Boyd, Jordan, &

Blackburn, 2015). *Appendix A* to this study lists LIWC2015 categories, sample LIWC dictionary words, word count in each category of the dictionary, and relevant scales.

LIWC's Versatility

Since its creation in the mid-1990s, LIWC has been continuously used in a large body of studies conducted in a variety of disciplines across a wide range of research areas. To include but a few, these studies relate to analyses of the prisoners' language for deception (Bond & Lee, 2005), interview differences between physicists and writers (Djikic, Oatley, & Peterson, 2006), the value of trauma essays for coping with trauma (Klest & Freyd, 2007), the differences in essays composed by normal and psychiatrically disturbed individuals (Junghaenel, Smyth, & Santner, 2008), self-narrative differences in personality (Hirsh & Peterson, 2009), and online narratives by female patients suffering from breast cancer (Bantum & Owen, 2009). This excludes a vast number of unpublished, doctoral dissertations that have applied LIWC.

LIWC for Gendered Language Analysis

The tool has also been widely used in studies for finding differences in language use by men and women in English and other languages. While some of these studies used select categories of LIWC, others employed all categories of various LIWC versions.

Using limited categories of LIWC, for instance, a study (Lester, 2014) compared suicide notes written by men and women and found that women were more concerned with themselves, the reason of their suicide and others. Men, in contrast, were found to be less personal and their notes contained more instructions. Another study on analysis of language of dreams for identifying male and female dreamers (Wong, Amini, & Koninck, 2016) found that of LIWC categories,

Psychological Processes (expressions pertaining to ingestion, contrast, perceptual processes, time, and motion) and personal concerns (expressions concerning death, achievement and leisure) were found to be significant indicators of the dreamers' genders.

Similarly, applying particular LIWC categories, another study (Yu, 2014) was carried out to analyse a huge corpus of speeches delivered at the US Congress from its 101st to 110th sessions (1989–2008), to see gender differences in language use in a formal setting of political discussions and debates. The study found that while both the genders adhered to a formal style (longer words and fewer pronouns), the language of women legislators showed language characteristics of both women (e.g. more use of emotion words, fewer articles) and men (e.g. more nouns and long words, fewer personal pronouns). Following a trend analysis, the study observed that gender differences in language use in Congressional speeches have been constant over the last 20 years, regardless of the topic of debate. The study argued that language differences in genders persist irrespective of the settings.

In contrast to this limited use of LIWC, a more comprehensive application of LIWC was made by studies that analysed language differences between men and women across all the LIWC categories. A relevant instance of this is the work undertaken by Newman, Groom, Handelman, & Pennebaker (2008). In this research, they studied a corpus of more than 14,000 texts from various academic disciplines across all 74 dimensions of LIWC2001 version. They found that women tended to use more words related to social and psychological processes. In contrast, men were found to refer more to impersonal topics and object properties. Although much of their findings supported previous research on language differences in genders, their study made an additional suggestion that difference of language use was larger in men and women on tasks that place fewer constraints on language use.

A more recent example of analysing texts across all dimensions of LIWC for exploring gendered language differences is the study of Schwartz, et al. (2013), who combined the techniques of closed vocabulary and open vocabulary analyses for a corpus of about 700 million words from Facebook messages. For the closed vocabulary analysis technique, they used all the 64 categories of LIWC2007. They claimed that many of their results aligned with the past research in language and gender studies. They reported, for instance, that females tended to use more emotion words and first person singulars and exemplify more psychological as well as social processes in their messaging. On the other hand, they found that males used more object references and swear words.

In the same vein, a study by Lester and Leenaarsb (2016) conducted text analysis of suicide notes by 513 men and 166 women across all 74 categories of LIWC2001 version. They found six substantial differences and four trends in both kinds of the texts. They concluded that, compared to the text written by men, women's notes had a higher percentage of words in the dictionary, present tense verbs, negations, cognitive process, words indicative, and discrepancies. They also observed that women's suicide notes tended to have language features indicating defeat-entrapment, hopelessness, and falling short of internalized standards.

Some of the adapted versions of LIWC in languages other than English were also applied to analysing language differences in genders. For instance, the Chinese version of LIWC—called CLIWC—was used by Lin, Lin, Wen, & Chu (2016) to analyse language difference of medical students in their reflective assignments. They studied the linguistic features in the psychological process categories of male and female medical students. They found cognitive words as more pervasive in comparison to perceptual words. They also concluded that female students exceeded male students in use of more positive words and words related to sadness whereas male students dominantly used words in the category of space.

Similarly, using the Russian version of LIWC, more recently Seredin & Lyell (2017) analysed a corpus of 113 male and female respondents on the topic "How I Spent Yesterday" for author deception detection in Russian narratives. The corpora were processed for seeing a significance between truthful and false Russian narratives. Claiming an accuracy of 68.3%, well above the pre-determined threshold of 50%, the study concluded that both men and women tended to tell lies.

2.4 Genre: Web Linguistics and CMC

The term "web linguistics" was first used by Bergh (2005, p. 45) to refer to empirical linguistic studies based on language data gathered from the Web. As a potent language source for CL research, Web is an enormous freely available source of language that can be subjected to all types of language analysis (Kilgarriff & Grefenstette, 2003). In Web linguistics, two main approaches are followed for use of online data: Web for Corpus (WfC), also called as Corpora from the Web, and Web as Corpus (WaC). The former is related to corpus construction from the Web data while the latter involves direct use of the Web as a corpus. The two concepts are often confused and WaC is used to indicate all aspects of web linguistics research even if they actually fall within WfC research (Bergh & Zanchetta, 2008).

In retrieving information from the Web, there are different search strategies, which can be categorised metaphorically into "hunting", "grazing" and "browsing" (Hawkins, 1996). In terms of linguistics, *hunting* is equivalent to seeking language elements directly from the Web, i.e. checking for a particular word or phrase directly on the Internet using search engine(s). In contrast, *grazing* means gathering language data sets that are already fabricated and maintained on the Web by some information provider, e.g. systematically downloading texts from some selected websites.

Finally, *browsing* may be explained as finding some relevant data by chance, e.g. finding and collecting subject and verb disagreement examples while randomly searching the Web. While each of these information retrieval strategies may be used for WaC and WfC research, there appears to be some correlation as *grazing* strategies are typically employed by WfC research while *hunting* strategies are used in WaC projects (Fletcher, 2007).

The advent of web-based communication has made more and more corpora accessible for linguistics studies. Additionally, the World Wide Web (WWW) not only offers a substantial amount of naturally occurring language data for CL research pertaining to the already existing genres or text types but also provides for research newly emerging document types and web-genres like wikis, blogs, personal home pages, commercial websites and so on (Mehler & Gleim, 2005; Thelwall & Wouters, 2005).

2.4.1 Weblogs

The origin of the term *weblog*, or its shortened form *blog*, is generally attributed to the weblog editor Jorn Barger who applied it in 1997 to describe his weblog *Robot Wisdom*, a collection of links to other web-based material (Blood, 2004). While Jorn Barger may be regarded as a pioneer of edited blogs, Justin Hall is considered the founding father of personal blogs, who started his own website (www.Links.net) as a student at Swarthmore in 1994 (Rosen, 2004). Early blogs were more like personal web pages of bloggers with links to related web contents.

A significant change to blogs occurred in 2002 when blogs were added with the feature of *comments*. After integration of this feature, blogs looked akin to bulletin boards, forums or news groups (Garden, 2011). With additional changes introduced by weblog designing software, many structural and functional features were added to blogs. These developments not only increased the

popularity of blogs and contributed to their rapid growth but also made their structure more complex, causing problems to categorise and define various types of blogs in clear terms. The major problem in its clear definition is that the term used for this CMC forum refers to it both as a technological platform and as an output. Therefore, current definitions of *blog* fall within two domains: one referring to its technical aspects; the other to its functional aspects. The former refers to blogs in terms of affordances and design features; the latter looks at them as a genre of communication (Scheidt, 2009; Treem & Thomas, 2010).

While detailed analysis of blog categories and its definitional problems may be found elsewhere (Garden, 2011), there is an overall consensus among scholars that a blog is a kind of web page or website designed and developed through a blog software for simplicity in creation and maintenance of blog contents. Such computer programmes provide a web-publishing interface and eliminate the requirements of high technical knowledge on the part of bloggers to upload contents. The format of a blog typically contains some basic units like the title of the post, the timestamp, the contents or the post, and the permalink (URL). The format sometimes also includes the name of a single author of a post, or names of multiple authors of a single post (Garden, 2011). There is also an agreement that blogs are generated and frequently updated by users. Blogs may contain texts, pictures or illustrations, and videos on which individuals, groups or communities share their thoughts, activities or interpretations. The entries of blogs (also called posts) are dated chronologically—the latest appearing at the top of the blog page—and archived. Readers generally have access to respond to a blog post (Eveland & Dylko, 2007). Blogs are distinct from other CMC in at least three ways: they foster personal individualistic contents, offer an easy and instantaneous publishing on the web and provide linkages to fellow bloggers to help build peer groups as found in the non-virtual world.

These features make blogs immensely popular among men and women nowadays. Even though blogs have gained popularity since 1990s, there has still been a growing number of blogs on the Internet. The blog search engine Blogpulse claimed to have counted about 152 million blogs on the Internet in 2013 (Johnson, 2017). Similarly, Kendal (2017) claims that, in 2017 alone, there were about 144,757 new blogs added with an average addition of about 1392 blogs per day. In 2017, Tumblr—a Yahoo-owned microblogging website—also claimed to have more than 373 million blog accounts (Statistica, 2017). As per estimates of Worldometres (2017)—an online calculator of daily blog based on statistics of Technocrati—there are more than 2 million blog posts published a day in the world.

Owing to the huge and ever-growing number of blogs around the world and the flexibility in its form and use, it is hard to create a fixed taxonomy of blogs. Blogs can, therefore, be a personal diary or a journal, a commentary platform, a news service, or a collection of random thoughts. However, some attempts have been made to categorise blogs. For instance, Hewitt (2005) divides blogs into two broad categories along a continuum from *pure aggregators* to *pure analysists*: The former indicates key links to readers while the latter analyses critical events and issues. Similarly, Hartelius (2005) differentiates between *group blogs* and *personal blogs*. The former is a blog typically published by a company, institution, agency or organisation with a multitude of contributors and in part of the organisation's website. Contrarily, the latter has one primary contributor and can be compared to personal diary or journal.

2.4.2 Newspaper Blogs

One sub-category of group blogs is the newspaper blogs. It was as early as 1993 that executive bodies of the newspapers started evaluating the use of Internet for the industry. While

some looked at it as a threat to their print-version newspaper industry, others saw it a great opportunity linked with technology. To address this dichotomy in opinions, the Newspaper Society, one of the world's biggest publishing association, funded a research project from levying members. Although the reports confirmed some of the threats, they tended to forecast significant opportunities for newspaper industry, especially the advantage of trusted newspaper brands over other media on the Internet (Beamish, 1998).

As a result of such preliminary efforts on blending the newspaper and the Internet, the first fully online newspaper *The Palo Alto Weekly* appeared in 1994 in USA (Carlson, 2009) and soon other newspapers in the country followed suit within a period of 18 months in the mid-1990s (Hall, 2001). These early developments involved a set of challenges to online newspapers as regards their design, management, and popularity (McAdams, 1995). Later studies confirmed the success of news industry by integrating social media and significant increase in its popularity through Twitter and blogs (Yahr, 2008; Farhi, 2009; Gleason, 2010).

Newspaper blogs provide a lot of audience interaction to newspapers that was not previously possible. This interaction is not limited to audience members only but also between journalists, editors and the audience. Newspaper blog consumers feel free to come up for an intellectual conversation on a topic of their interest that appears in the news. However, such conversations are monitored by newspaper blog editors who act as moderators and supervise discussion, which makes these blogs a healthy addition to news-related websites (Everett, 2011).

2.4.3 Language, Gender and Newspaper Blogs

An important notion within the context of technology, mass media and public discourse is the "patriarchal notion of public/private dichotomy associated with male/female" (Harp &

Tremayne, 2006, p. 249). This means that the world of women is largely limited to domestic environment whereas men are considered as best suited for public sphere. Based on this dichotomy, different social roles are given to the two genders. This notion of public/private dichotomy is extended to the roles of the two genders not only in the journalist practice but also in the Internet use (Gao & Martin-Kratzer, 2011). Studies suggest that genders in online communities are likely to strengthen the pre-existing gender patterns (Kramarae & Taylor, 1993). For this purpose, gender and CMC studies have applied the theories of face-to-face communication such as Lakoff (1975), Tannen (1990a, 1990b), and Coates (1993).

For example, in an analysis of blogs by men and women, research has shown that men were more interested in blogging about politics, external events and technology whereas women tended to blog about personal matters (Herring, Kouper, Scheidt, & Wright, 2004). In the field of journalism and news blogs, male bloggers have been found to include more male perspectives and male values like self-orientation, assertiveness, and independence while female bloggers have been reported to emphasise more on interdependence and cooperation (Armstrong, Wood, & Nelson, 2006).

While there is some research on language differences in men and women bloggers in general, there is extremely limited research found in analysing language differences in men and women bloggers of online newspapers. In the context of Pakistan, a single published study (Amjad & Rasul, 2017) that analysed blogs of online newspapers was only recently conducted. However, this research focused on portrayal of Pakistani working women in these blogs.

The scarcity of research on the language of newspaper blogs opens up a new avenue for research to explore the language features of this untapped online resource. The present study is, therefore, an attempt to venture into this area.

2.5 Context: Pakistan

2.5.1 Language and Gender Landscape

Pakistan came into existence on August 14, 1947 on the basis of ideological foundations of Islam. The country has four provinces: in alphabetic order, they are Balochistan, Khyber Pakhtunkhwa, Punjab and Sindh, besides its capital territory, Islamabad. Each of these areas has a unique ethno-linguistic background. The country has both urban and rural areas; the former is much developed but the latter still lacks many facilities.

Table 2.1

Sex ratio: Population in Pakistan

Area	Male	Female	Transgender	Total	Male/Female Ratio
Rural	67,300,171	64,886,593	2,767	132,189,531	103.72
Urban	39,149,151	36,428,187	7,651	75,584,989	107.47
G. Total	106,449,322	101,314,780	10,418	207,774,520	105.07

The demographic data released by Bureau of Statistics (2017) shows Pakistan as a country populated by 207.774 million people. Muslims form majority in the country with about 96% population. As given in Table 2.1, the sex ratio of population in Pakistan, based on Bureau of Statistics (2017), shows that men, women and transgender form 51 and 49 and .005 per cent of the country's population respectively. Of these, 63% men, 64% women and 27% transgender live in the rural whereas 37% men, 36% women and 73% transgender live in the urban areas.

The linguistic landscape of Pakistan is extremely rich and imbued with different colours. There are about 72 languages spoken in Pakistan (Rahman, 2010). Of these, English is one of the official languages in the country alongside the national language, Urdu. English is taught to Pakistani leaners on the principles of English as a Foreign Language (EFL). Most of Pakistanis are multilingual (Ahmed, Ali, & Xianfa, 2015) who acquire their mother tongues (e.g. Pashto, Punjabi, Sindhi, Baluchi, or Siraiki, which remains mostly their first language) at home while they learn

Urdu and English at schools where both are taught as compulsory subjects. Being the national language, Urdu is widely understood and used for communication in the country. Though English is taught in all public- and private-sector schools in Pakistan, it is used as a medium of communication more than Urdu in the urban areas of Pakistan, that too at upper socio-economic levels (Rahman, 2010). One possible explanation for this could be the socio-economic prestige associated with English in the country (Rahman, 1999). Since the creation of Pakistan, English has widely been used in public sector institutions and urbanised areas of the country. As it is a popular medium of communication in offline contexts, it is also preferred over vernaculars in Pakistan for various communicative and functional purposes over the Internet, including blogging.

2.5.2 Language and Gender Research

Research on gender and language has become an obvious, universal and frequently observable phenomenon both at national and international levels. Although the number of studies in language and gender has grown considerably elsewhere, there is limited work done in Pakistan to fully explore multiple aspects of this relationship in various genres and contexts. The present study does not claim to have reviewed all the language and gender studies conducted in Pakistan. It has, at least, accessed a reasonable part of the published works over the last two decades.

In the multilingual linguistic landscape of Pakistan, the rich interplay of languages has its manifestations in gendered language use in bilingual and monolingual contexts, both offline and online. Unless otherwise specified, bilingual context in this sub-section of the review (i.e. 2.5.2 Language and Gender Research) means English-Urdu interaction whereas monolingual context means studies conducted on communication in English.

In bilingual context, a study (Hamid & Loewenthal, 1996) was conducted to infer a writer's gender by studying gender differences in handwriting of men and women in Urdu and English.

The study involved 30 donors (14 men and 16 women) whose handwriting was judged by 25 judges. Claiming a judgement accuracy level of 68%, the study concluded that in Urdu the writing of women was more decorative and delicate compared to the handwriting of men. However, the study did not report any significant difference in the handwriting of men and women in English. In the same context, another study (Ali & Aslam, 2012) examined the effect of learned words on gender in the language used in 1000 English short messaging service (SMS) texts in Pakistani society where Roman Urdu is also used as a language for mobile phone communication. The study collected 1000 SMS texts from 25 male and 25 female university students of Lahore and reported that female used learned words of English more frequently than male. It also found that female used SMS more than male to communicate. In SMS writing, a more recent study (Ahmed, Ali, & Xianfa, 2015) investigated gender-based code-mixing in Pakistani society. The data was gathered from students of different universities in Lahore. Through random sampling, 1000 messages of 50 students (25 male and 25 female) were analysed. It found that the messages of females contained more words in comparison to males and that females code-mixed more than their male counterparts. The study further observed that females code-mix when they discuss topics pertaining to entertainment, education, and personal matters and that they generally stick to one topic. Males, on the other hand, code-mix less frequently and discuss a variety of topics. Besides, females code-mix to make their message convenient, save time, and create fun while males do so to ensure understanding of expression, express clearly, and use the right expression. Based on these findings, the research concluded that, in terms of code-mixing, there exist clear gender-based differences in SMS language in Pakistan. Similarly, research on investigation of bilingualism and gender identity construction in online discourse on Facebook (Rafi, 2017) collected data from 200 students from 5 different universities of Lahore. It observed that male and female users of

Facebook differed in their choices of drawing gender boundaries in their language. It found that females tended to be more innovative and used expressions that were of foreign origin, e.g. dropping of 'd', 'g', and 't'. In contrast, male users were found to be more frequent users of English code-switching, especially when they communicated with females to express power or show a serious attitude.

In monolingual contexts, while some studies were conducted on differences in language use by men and women in intra-gender communication, others saw this relationship in inter-gender communication. For instance, a case study in intra-gender conversation (Gul, 2010) described views on gender-based speech variations in the working environment of the Pakistan Air Force. The study combined a questionnaire with interview to analyse the particular views. While the study confirmed many of the previous research findings, it refuted others. Similarly, another study (Shah, Mahmood, & Qureshi, 2011) explored variations in language use due to gender-based social constraints in Punjabi society. It administered a questionnaire based on four main questions of the study to a group of 24 college and university teachers (12 men and 12 women). The study analysed the views of the respondents and reported that in Punjabi society women tend to use more polite language, the standard variety of Punjabi, and more declarative than imperative and assertive forms in addition to concluding that the social constraints affect the use of language. It further noted that men also use more declarative form alongside assertive and imperative forms. In the same manner, another study (Kousar, 2013) examined speech acts phenomenon in Pakistani context by observing realisation patterns of apologies in Urdu across variables of severity of offence, social difference, social status, and gender. The research selected a sample of 152 students from four cities namely, Lahore, Sheikhupura, Shahkot and Faisalabad of the Punjab province. It concluded that the variable of gender did not influence the choice of the apology. However, gender differences were

observed in the preference for the strategies: men were found to acknowledge their responsibility for the transgression and apologised openly. Women, in contrast, were found to give explanations more and offered compensation or repair for the loss.

In inter-gender conversation analysis, some studies (e.g. Noreen & Zubair, 2012) were pinned in the anti-essentialist approach, which analysed leisure talk recordings of eight conversations between close female friends. The participants were in their early twenties and each recording lasted for 60 minutes. The analysis focused on femininity construction on the basis of linguistic patterns of overlapping, tags, hedges, intensifiers, compliments, repetition, latching, and backchannel support. On the basis of its analysis, the study challenged the binary classification of gender as advocated by the *deficit* and *dominance* approaches and supported the claim of the *dynamic* approach—though without naming it—that gender as a variable is located in local interaction. The study concluded that instead of showing their weakness of powerlessness, these speech patterns indicate the tendency towards intimacy, cooperation, support, involvement and connectivity among women.

While these studies analysed spoken language, others were conducted to see gendered language differences in written language. An instance of this is a relatively recent study of SMS writing by men and women (Rafi, 2008), which examined two assumptions: SMS language is like a pidgin in communities; and, females have greater motor of SMS and choose different lexical and morpho-syntactic choices than males. The study collected 100 text messages randomly from 20 phones and analysed perceptions of 25 male and 25 female users. It analysed the selected texts at lexicological, morphological and syntactic levels and concluded that there is a significant difference between males and females in lexical and morpho-syntactical choices. It, however, found no significant difference in the perceptions of the genders about the impact of SMS on the

language used in commercials. It also found that females were more skilful in writing longer, more complex and lexically more dense messages in comparison to males. Later, gendered SMS texts were investigated against the variables of compression, abbreviations, symbols, tenses and punctuation to find out gender boundaries (Rafi, 2010). The research concluded that SMS text identifies some gender boundaries: for instance, females tend to use more compressed forms of words, abbreviations, and acronyms than males. Moreover, it found that there is a significant difference between genders in the use of standard grammatical structures and punctuation. For example, females adhere to more norms and standards than males. A study on use of catchy words in a mix of spoken and written texts (Rustam, 2010) analysed differences in use of 15 selected catchy words (like 'chill pill', 'lash pash', 'fit fat'), which have no standardised meaning. It gathered data from college and university level students and analysed it across different variables including medium of communication and gender. The results showed that female students tended to use catchy words more than the male students. It also found that catchy words were used more in SMS compared to its use in face-to-face communication, letters and telephone calls. A more recent attempt on exploring gender differences in word formation processes (WFPs) in SMS language (Zaheer, 2017) gathered data from 50 male and 50 female graduate-level students of Business Administration from a Pakistani university and found that males and females follow different WFPs, both standard and nonstandard. It also reported that females tend to use more WFPs like acronyms, clippings, homophones in comparison to males and that females did not use standard forms of language in their SMS writing. Based on these findings, the research concluded that WFPs could mark gender identities in SMS.

In online communication, differences and similarities between men and women Facebook users were traced (Nazir, 2012). The study examined the way genders use Facebook besides

highlighting their language differences. It selected 60 profiles, 30 each for men and women over a period of one month and found that women go for groups that relate to studies and fashion while men go for groups that are related to business. Moreover, in choice of topics for discussion, women were found to write more about their emotions, which they convey through the vehicle of poetry, examinations, psychology, studies, fashion, and weather. In contrast, men were found to write more about business, politics and sports. Women were also found to use more emoticons to maintain face and keep conversation going on. The study, however, did not find any significant difference in the use of Internet language by the genders. In the same vein, a corpus-based study (Naveed, 2014) analysed cyber language to identify gender boundaries among one hundred Facebook users of various age groups (13 to 30 years). The results indicated that a new but easily understandable language has evolved through Facebook, which is responsible for significant differences between male and female linguistic properties. A latest study on online language (Kamran & Mansoor, 2017) compared the written comments on Facebook by 220 Pakistani male (62%) and female (38%) students of private- and public-sector universities. While the study found some regional differences in the representation of male and female students on Facebook, it also highlighted that, compared to male students, female students used more emotional language in their comments. It also found that female students tended to post more comments about fashion, women's progress, diet, nostalgia and party arrangements.

A critical analysis of these studies reveals that research on differences in language by male and female genders in Pakistan is a growing phenomenon in the country context. However, this research so far has studied either conversation analysis or analysis of language used in SMS and Facebook. One significant area so far underrepresented in this research is Pakistani blogosphere. The existing research on blogs in Pakistan is not only limited in scope but also carried out on topics

such as female gender portrayal in newspaper blogs in language and gender research (Amjad & Rasul, 2017) and developing a new computer algorithm for sentence-based semantic analysis of blogs in Pakistan in the field of artificial intelligence (e.g. Aziz & Rafi, 2010). Thus, Pakistani blogosphere is a huge resource yet to be fully explored by studies analysing gender differences in language use. Additionally, the existing research is based on limited data and small samples taken from particular institutions and cities, which cannot be representative of the entire population in Pakistan. Also, majority of the studies have used traditional content analysis procedures and applied manual analysis techniques, which lack scientific rigour and are prone to subjectivity. Some of the studies have not analysed the actual language used by men and women but views of study respondents on use and features of gendered language.

2.5.3 Pakistani Blogosphere

The Internet facility plays a key role in the popularity of blogs in Pakistan. The facility in the country is much better now than it was in the past. Since 2011, the government has approved a 3G network in the country, which has increased Internet use on mobile phones in addition to other computer devices. However, Pakistan is still not sufficiently equipped with this facility. Only 16% of the total population has access to the Internet. This figure, however, does not include the 29 million broadband subscribers mostly residing in the urban areas (Yusuf, 2013).

In Pakistan, the Internet is growing in size and influence. Every year, more than a million persons in the country sign up to social media networks (Rafi, 2017). They are mainly motivated by the potential of these networks for staying connected to one another and sharing, what they think, is important for them. Social networks not only enhance a sense of community or shared identity in members but also afford them the freedom to express their thoughts and feelings with a much larger community available online.

Expansion in the Internet has increased trafficking on all Pakistani social media networks, not least the Pakistani blogs. In Pakistan, almost all electronic media channels have their websites and blogs. However, perhaps due to limited Internet facilities, the rural areas of Pakistan are still far behind to catch up with the growing increase in the blogging facilities.

Established media groups in Pakistan increasingly seek to attract online consumers through their own blogs. Therefore, besides a growing number of independent blogs, "the blogs of established media groups dominate the Pakistani blogosphere" (Yusuf, 2013, p. 8). Both leading Urdu newspapers like *Daily Jang* and English dailies like D, TDT, TExT, TNn, and TNw publish blogs. Besides publishing blog posts of their journalists and staff members, the online English newspapers, called e-newspapers, maintain rosters of a number of independent bloggers who like to publish their posts on these established e-newspaper websites for greater exposure and better remuneration (Yusuf, 2013). Table 2.2 presents an outline of the English e-newspapers whose blogs have been chosen for study in the present research. The list has been arranged chronologically from the dates of the origin of the print version of the newspapers.

Table 2.2

Outline introduction of English e-newspapers chosen for the study

S. No	Name	Date launched	Founder	Place(s) of Origin
1	D	Oct 26, 1942	Muhammad Ali Jinnah	New Delhi
2	TNn	Oct 1, 1986	The Nawa-i-Waqt Group	Lahore
3	TNw	Feb 11, 1991	Jang Group	Karachi; Lahore; Rawalpindi; London
4	TDT	Apr 9, 2002	Mr. Salman Taseer	Islamabad, Karachi , Lahore
5	TExT	Apr 12, 2010	Lakson Group	Karachi

2.6 Gap in the Present Research

The above discussion shows that there exists a wide gap to be filled by studies in the field of language and gender in general and in the context of Pakistan in particular. This gap can be identified at all strata mentioned here under different sections: *theory*, *method*, *genre* and *context*.

At theoretical level, there are several reasons for conducting further research. First, the debate among the *essentialist* and *social constructionist* theories is still going on, especially in the wake of recent debates in the field of anthropology and evolutionary psychology that gave birth to what has been termed as "new biologism" (Cameron, 2009, p. 2)—a theoretical belief that re-emphasises attribution of gender differences to biological sex. Therefore, there is a need to explore further as the final consensus in academia on relationship between language and gender has yet to be reached. Second, most of the research conducted on language and gender has taken place outside Pakistan. With its own socio-cultural make-up, Pakistan offers a rich field for social scientists and linguists to explore multiple dimensions of gendered language phenomenon. Within Pakistan, much of the research on language and gender focuses on functional aspect of language, i.e. language as a vehicle for gender identity, gender supremacy or gender discourse construction, not the specific gendered linguistics features.

At methodological level, previous research has predominantly used traditional linguistics research methods with limited data, which was manually coded and analysed by linguists. Since CL research is a relatively new phenomenon, not many studies have been carried out to see the relationship between language and gender. In Pakistan, this situation is even graver as CL studies on this relationship are extremely rare.

Moreover, the study can contribute to research pertaining to the particular web-genre. The advent of the Internet has produced many web-based text genres, weblog being one of them. Previous research has focused more on face-to-face conversation analysis or offline written texts. The vast repository of online texts is a new resource for linguists to explore. As this review shows, there is limited research into language differences between males and females in CMC, not least

the weblogs. Within weblogs, the sub-category of newspaper blogs is yet another sub-genre that needs to be explored.

The context of Pakistan offers its own particular socio-linguistics landscape. The country context makes the present study unique in all three aspects—theory, method, and genre—highlighted above. To the best of the researcher's knowledge, there has, so far, been no previous research carried out to study language differences (theory) using CL research techniques (method) to explore e-newspaper blogs (genre) in Pakistan (context).

CHAPTER 3: DATA & METHODS

3.1 Chapter Overview

To fill the research gap highlighted in the previous chapter, the researcher developed GenCorB of leading Pakistan English e-newspaper blogs and analysed it through CL techniques, as explained in this chapter. Some important considerations were kept in mind while building and analysing GenCorB. The chapter has thus been divided into four main sections with different sub-sections. The first section deals with corpus design. The second section explains data collection procedures, i.e. building GenCorB, which highlights the need for the corpus and its overall make-up in terms of data source, data preparation, representation and size. The third section explains CL methods applied to analyse GenCorB and deals with particular corpus analysis software and techniques applied in this study. The final section covers additional research considerations like copyright issues and research ethics pertaining to this study.

3.2 Corpus Design

In today's digital, fast-progressing and technology-intensive world, CMC has emerged as a new form of language (Baker, 2010a). As part of this digital boom, digital journalism (i.e. communicating along with discussing news and events round the globe through social networking websites, e-newspapers and blogging) is gradually becoming the need of the day (Kaukab & Mehrunnisa, 2014). E-newspaper blogs are increasingly used for effective communication of views on a vast array of topics by both male and female bloggers. They are, therefore, expected to contain rich linguistic resource for studying differences in language use by men and women. Since CMC is a relatively new filed, not all CMC genres are currently represented in large balanced

general corpora. Resultingly, corpora for CMC genre are often *specialised* in nature and developed for meeting specific project requirements (Beißwenger & Storrer, 2008).

As Hunston (2008, p.155) argues, "It is a truism that there is no such thing as a 'good' or a 'bad' corpus, because how a corpus is designed depends on what kind of corpus it is and how it is going to be used." Since the study aimed to analyse the language of Pakistani men and women, the variables involved in the design criterion were gender, age (i.e. adulthood) and nationality (Pakistani). Two specialised gendered corpora were designed for comparison. The texts were selected on an external criterion. The blogs were available on official websites of the five e-newspapers. The corpora were specialised in nature, intended for use by the present research only. The issues of representativeness, balance and size were taken into consideration as explained in the succeeding sections. Full texts of blogs were included in the two corpora to cater for rare features occurring in language, which often require larger text samples.

The population from which GenCorB was sampled was determined on the basis of language production (McEnery, Xiao, & Tono, 2006). For such a population, no sampling frame already existed, which could have facilitated *probability sampling*. Therefore, non-probability (*purposive*) sampling was applied and appropriate cases were selected for study.

The overall design of GenCorB corresponds to Type 1 Corpora in CMC as classified by Beißwenger and Storrer (2008). Within CMC genre, GenCorB was built on the model of another specialised corpus *The Blog Authorship Corpus* by Schler, Koppel, Argamon, and Pennebaker, (2006). However, unlike *The Blog Authorship Corpus,* which focuses on the role of age and gender on blogging practices, GenCorB focuses on the effect of gender on language only. Other typological aspects of the research design for the present study have been summarised in Table 3.1 with a check mark next to the related dimensions.

Table 3.1 Typology of GenCorB	
1. Overall Research Design (Brezina, 2018, pp. 21-22)	
Whole Corpus Design	-
Individual text/Speaker Design	✓
Linguistics Feature Design	-
2. Definition of Population (McEnery, Xiao, & Tono, 2006)	
Language Production	✓
Language Reception	-
Language as a Product	-
3. Corpus Structure (Horváth, 2006, p.113)	
By Language	
Monolingual	✓
Parallel	-
L1	-
Learner	-
By Representation	
Synchronic	✓
Diachronic	-
General	-
Specialised	✓
By Text Type	
Written	✓
Spoken	-
Combined	-
By Storage	
Static	✓
Dynamic	-
By Notation	
Un-annotated	✓
Annotated	-
By Status	
Set	✓
Developing	-
By Use	
Linguistic	-
Applied Linguistic	✓

3.3 Data: Building GenCorB

3.3.1 Why to build GenCorB?

While there are sizable general and specialised corpora available, they were not found suitable for the present research for the reasons explained in this section. General corpora like the BNC could not be designed/built for the present study for two main reasons. The first reason is related to the aim of the research, i.e. general corpora are compiled to represent language as a whole with a variety of genre representation and balance. Therefore, their purpose is to infer generalisations about the language as a whole (Aston, 2001) and not to research a special language use, e.g. the way men and women write blogs, as is the case in the present study. The second reason was of methodological nature: although some *general* corpora such as the BNC were available online, they were not *specialised* corpora of the target language for this study, i.e. language used in blog posts.

A *specialised* corpus *The Blog Authorship Corpus* of blog posts containing over 140 million words was also downloadable from http://https://u.cs.biu.ac.il/~koppel/BlogCorpus .htm but it was not used for the present study for three reasons: The first one was related to the specific variety of the language in question. The text of the blog posts of *The Blog Authorship Corpus* did not represent the Pakistani variety of English. The second reason was based on the context of the language. The text of the said corpus was not produced in the socio-cultural context of Pakistan. Since language cannot be separated from its context (Widdowson, 1998), this is where the value of and need for a *specialised* corpus representing a particular language context becomes all the more obvious. One aspect of the *specialised* corpora value is that the analyst of the corpora is mostly their compiler who is familiar with the wider socio-cultural context in which the language is produced and who acts as the "ethnographic specialist informant to shed light on the data"

(Flowerdew, 2004, p.16). Finally, the text was not gathered from the specific sub-type of blogs, i.e. the Pakistani e-newspaper blogs. For these practical considerations, a need was felt to build GenCorB. The building of GenCorB, which has been explained in a step-by-step manner in the subsequent part of this chapter.

3.3.2 Data Source

The study required building two *specialised* comparable gendered corpora: one comprising texts of men bloggers (MenCorB), the other of women bloggers (WenCorB). The data required for the study was collected from the blog posts of five leading English daily e-newspapers of Pakistan. In an alphabetical order, these newspapers are *Dawn* (D), *The Daily Times* (TDT), *The Express Tribune* (TExT), *The Nation* (TNn), and *The News* (TNw). There are other reputable English daily e-newspapers in the country like (alphabetically) *Pakistan Observer*, *The Frontier Post*, *The Post* and *The Statesman*, but they do not maintain blogs so far. From these selected newspapers, data for the present study was gathered from blog posts between November 01, 2008 (the date of the oldest selected blog posted on an e-newspaper) to August 31, 2015 (the culmination date of the researcher's PhD coursework phase, which followed the beginning of his research phase). The data did not get a representative sample of the selected bloggers' posts in the form of segments but included full posts, which were retrieved from the e-newspaper blog archives during the said timeframe. Table 3.2 represents the dates of the oldest blog posts in each newspaper that were included in this study.

Table 3.2

e-Newspaper-wise dates of oldest blog posts

S. No	e-Newspaper	Abbreviation	Date of Oldest Post
1	Dawn	D	November 01, 2008
2	The Express Tribune	TExT	January 01, 2010
3	The News	TNw	September 24, 2011
4	The Nation	TNn	July 03, 2013
5	The Daily Times	TDT	April 01, 2015

The blog posts of these e-newspapers were available in two separate ways of archival arrangement: Four of the e-newspapers, i.e. D, TDT, TExT, and TNn arranged their blog archives in a date-wise chronological order, while TNw archived blog posts in two ways, i.e. under the bloggers' list and under pre-defined categories. This archival arrangement of blog posts made it relatively easier for the researcher to access, retrieve and store data.

3.3.3 Data Collection and Preparation

The process of data collection and preparation for the study involved three phases. The first phase comprised manual collection and storage of all blog posts from the archives in MS Word file format. The decision to build the gendered corpora by extracting and storing all blog posts manually was taken for two reasons: The first was of technical nature, i.e. the lack of knowledge and expertise on the part of the researcher of any available software (WebCrawler or bot) that could have been used for automatic data mining and storage from the e-newspapers blog archives. The second reason was practical and more serious in nature, i.e. to ensure that the gendered corpora are representative of the target population, not all blog posts could be included in the study even if they could have been retrieved through a bot.

After having accessed and stored the blog posts manually, each blog post was extracted and stored as a separate MS Word file under a filename showing the blogger's own name preceded by a mark for gender. The abbreviated name of the e-newspaper followed by the letter MF (merged files) were added to the end of the file name for easy identification. For instance, the made-up file names 'M – ABCDEFG TExT MF' and 'F – HIJKLMN TNn MF' mean that the first file contains a text of a male (M) blogger whose name is ABCEFG, the e-newspaper is *The Express Tribune* and this file contains all merged blogs posts of the author; the name of the second file indicates that the text in the file was written by a female (F) blogger named 'HIJKLMN', whose blog posts from e-newspapers *The Nation* have been merged. In the same way, all the selected gendered blog posts were gathered from each e-newspaper and stored in separate folders of MenCorB and WenCorB. In all, 1674 files for men bloggers and 1212 files for female bloggers were retrieved and stored in MS Word format (.doc).

After having manually downloaded and stored the data, the second phase was to clean the data of any noise. Noise in real life data is a reality and it is a common feature of more informal texts like blog posts (Agarwal, Godbole, & Punjani, 2007), which has been defined as "any kind of difference in the surface form of an electronic text from the intended, correct or original text" (Subramaniam, Roy, Faruquie, & Negi, 2009, p. 115). It may be a result of conversion of text into digital form (e.g. conversion of printed/handwritten documents using OCR) or when the text originated in digital form (e.g. in informal settings of emails, chats, SMS, blogs). Also, what may be considered as *noise* for automated processing of text may not be considered so if the text is processed manually and *vice versa* because humans and machines make sense of texts differently (Subramaniam, Roy, Faruquie, & Negi, 2009).

While removing noise from GenCorB, the phrase "intended, correct or original text" (Subramaniam, Roy, Faruquie, & Negi, 2009, p. 115) was kept in view. To ensure that only such texts are included in the corpora, the blog posts were thoroughly examined and cleaned. Noise in the present corpora was of two kinds: The first was of *external nature,* i.e. it did not appear to be originated by the blogger but seemed to be added to the data by the blogger or any other agent. This included clearing the data of any kind of sources or internet links, footnotes, editor's notes, pictures and cartoons, embedded videos, and advertisements. The second kind of noise was of *internal nature,* i.e. it appeared to be blogger-originated. This type of noise clearance included removing typological or formatting errors, text underlining, bullets, and extra dashes used as section separators within blog posts. In such a case, copying data directly from websites to clipboard usually copies the Meta tags, Google Ads scripts and some hidden HTML tags, which is undesirable. The copied clipboard data, when pasted directly into a plain text file editor (e.g. Notepad), put the undesirable data along with the real blog text. The clutter so created, needs to be removed manually from each blog post, which is a cumbersome task. The saving of files in MS Word format was of great help as the programme's inbuilt features (i.e. *Spelling & Grammar*) were used for clearing such noise.

Once all the blog posts from the e-newspapers were collected, stored and cleaned of noise in the manner explained above, the third phase began, which involved preparation of the data for preliminary computation. Since manual conversion of all 2886 MS Word (.doc) files (men's 1674 and women's 1212) one by one into Text Documents (.txt) was an extremely time-consuming and laborious task, a custom-built software named "FileAttributesReader" was developed with the help of a software engineer (see acknowledgement section of this thesis) to convert files from Microsoft Word format (.docx, .doc) into plain text file format (.txt).

The software is a customized package developed specifically to serve the desired purpose only. It was designed in C# language using .Net framework and used standard code libraries registered with .Net Platform. The success rate of conversion process was 100% with no data loss. Besides being capable of converting the MS Word (.doc) files into Text (.txt) files, the software was customised to extract, into MS Excel format, metadata on each file, e.g. filename, date of blog post, number of words in a post, and topic of the post. Completion of the three phases of data collection and preparation made the corpora building process rather slow.

3.3.4 Corpus Representativeness

During data collection, an important consideration was to make the corpora as representative of the target population as possible. *Representativeness* is a commonly accepted characteristic of any corpus, which distinguishes a *corpus*, a purposefully and systematically collected body of natural language representing some variety, from an *archive*—just a random collection of written or spoken texts (Xiao, 2010). Also, a corpus may be called as representative of a language variety if the results based on the corpus are generalisable to that language variety (Leech, 1991).

There are two criteria for achieving representativeness of a corpus: *external* and *internal* (Aston & Burnard, 1998; Sinclair, 2004) or "situational and linguistic perspectives" (Biber, 1993, p.243). During text selection for corpus representativeness, the former criterion takes into consideration non-linguistic features or characteristics external to the text, which distinguish one text from another, whereas the latter criterion looks into linguistic features of texts or characteristics internal to the text.

The criteria mentioned above typically apply to *general* or *monitor* corpora and may not hold good for *specialised* corpora (Sinclair, 2004). Representativeness of a corpus is an "act of faith" (Leech, 1991, p. 27) as there is no means to ensure and evaluate it objectively (Tognini-Bonelli, 2001). In case of *specialised* corpora, representativeness is compromised at the outset because of "the need to target a disciplinary or thematic speciality" (William, 2002, p.45) as well as because *specialised* corpora need to be built following an external selection criterion, which relates to *who* created the text, for *whom*, and *what* the subject matter of the text was related to (Flowerdew, 2004).

The present study employed the *external criterion* to achieve representatives of MenCorB and WenCorB. The decision was based on the intended use of the corpora. The corpora were built to study differences in the language use by adult men and women bloggers of Pakistani English e-newspapers, which, in turn, were dependent on studying the linguistic features of the gendered texts. So, if these linguistic features had been predetermined (*internal criterion*), there was no reason for analysing the corpora. As Sinclair (2004) argues, texts in a corpus should be selected through *external criterion* so that their linguistic features remain independent of the selection process.

In application of the *external criterion* for corpus representativeness, the definition of the target population is a foremost requirement, which depends upon two aspects (Biber, 1993). The first aspect is to determine the population boundaries, i.e. to decide which texts to include or exclude. The second aspect is the "hierarchical organisation within the population" (Biber, 1993, p. 243), i.e. to see which categories of texts are included and to define such categories.

For the present study, the target population was the adult male and female Pakistani bloggers of leading Pakistani English e-newspapers. Therefore, applying *judgmental sampling*

(also called as *purposive sampling*), the blog posts representative of the intended population were to be included in the study, which, in turn, required that the population boundaries be determined first. For this purpose, the language of blog (i.e. English) and blog type (i.e. Pakistani e-newspaper blogs) were kept in mind. Also, the bloggers' age, nationality, and gender were taken into consideration. To ensure the language of the blogs and the blog types, English e-newspapers of Pakistan were selected for the study. For the remaining population boundaries, i.e. age, nationality, and gender, the bloggers' profiles were a useful source.

Bloggers' profiles have been a dependable source for researchers to obtain information on variables like bloggers' names, gender, age, education, location and other necessary information as indicated by research studies. For instance, in a longitudinal study of 357 web blogs, Herring, Scheidt, Bonus, & Wright (2004) found that 92% of blog authors provided a name, whether it was a full name (31%), a first name (36%), or a pseudonym (29%); 94% their gender; and, 90% their age. Also, more than half (54%) of blog authors provided explicit demographic information such as occupation or geographic location. More specifically, in case of e-newspaper blogs, a newspaper "requires users to register before participating. This strategy keeps participants from posting anonymously. . ." (Gsell, 2009, para. 3).

The bloggers' profiles available for the present study were no exception to what these studies have found. For instance, in case of age, the information in most of the cases was not explicitly stated. So, it was obtained either from where such explicit information was available or inferred from related information that implicitly indicated the bloggers' age (e.g. information on their profession or level of education such as professor, businessman, or a BS student). Since the present study was of adult bloggers, the age of adulthood was considered to be 18 years as laid

down by Pakistani law, i.e. *The Majority Act 1875* (Government of Pakistan, 2006). Therefore, bloggers younger than 18 years were not included in the study.

The second parameter of the target population was to determine gender of the bloggers to build gendered corpora, which was by the bloggers' names and pictures given in their profile. In some cases, names were not clearly indicating the bloggers' genders; so, the accompanying pictures were of immense help. Also, English gendered pronouns (he/she) used in the bloggers' profiles were helpful in finding out the bloggers' gender. In this regard, due caution was exercised and bloggers with vague identity were excluded, e.g. those who used abbreviations instead of names without pictures or identified themselves with avatars without having any other explicit gender identity indication in their profiles. Blog posts jointly written by a male and a female blogger or by bloggers' groups were also excluded for the obvious reason of indefinite contribution to the text by either gender. Similarly, anonymous blog posts or those using vague aliases or initials, with no pictures, were also not included in the data.

Besides their age and gender, bloggers' nationality was another parameter for the present population. Since the study focused on the context of Pakistan and Pakistani variety of English, bloggers whose nationality was other than 'Pakistani' were not included as they did not represent the linguistic and the socio-cultural context of Pakistan.

Once the population boundaries were determined, the second aspect of the population definition was considered, i.e. to see which text category has been included in the data. Broadly speaking, the text category selected for the present study belonged to a particular web genre, i.e. blog. Within this broader category, the texts were selected from the Pakistani e-newspaper blogosphere, which belongs to the category of *group blogs*. The reason behind choosing texts from e-newspaper blogs was twofold: First, in newspaper blogs, bloggers feel freedom to enter into an

intellectual conversation on any topic that appears in the news or is of social significance. This freedom of expression, in turn, provides an opportunity to linguists to analyse the language in a naturally expressed way. Second, in Pakistan, blogs by leading media groups dominate the country's blogosphere (Yusuf, 2013). Therefore, these blogs provide a huge and varied linguistic resource to researchers as bloggers from different professional backgrounds contribute their posts.

For the present study, information on the professional background of bloggers was obtained from the bloggers' profile. However, not in all cases were the bloggers' professions explicitly mentioned in these profiles. While some of the profiles clearly contained this information (e.g. student; doctor; lawyer), others mentioned it in general terms (e.g. social worker; political activist) whereas some others did not contain such information at all.

Similarly, bloggers from some of the professions had a reasonable size in the population while others had less representation. The number of bloggers from homogeneous professions with less representation were counted together to afford them visibility in the data. Table 3.3 explains who were grouped in these categories for counting. For ease of data handling, the names of the categories have been abbreviated and these abbreviations have been shown in the table. Figure 3.1 plots on a clustered column chart the representation of these professions in GenCorB and shows that information about professional background of the bloggers was available in 89% cases. In about 2% cases, the information was not available whereas, in about 9% cases, the profession was not clearly mentioned. In the remaining 89%, students (24%) and journalists (22.6%) are major contributors who were included in GenCorB. The data obtained from the profiles show that bloggers included in this study came from diverse professional and educational backgrounds, which provided diversity to the themes discussed in their writings and variety to their language use.

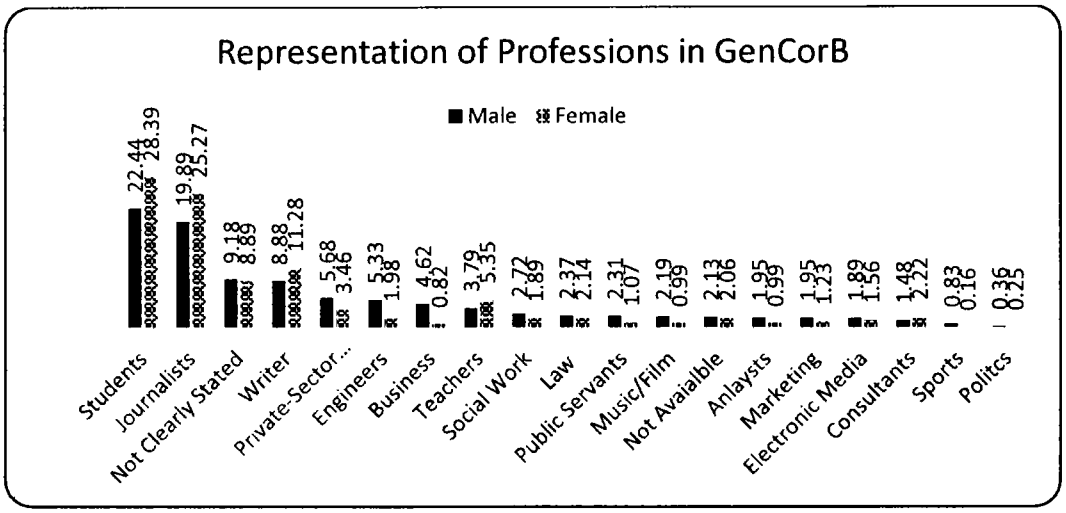


Figure 3.1. Percentage of bloggers' professions in GenCorB

Table 3.3

Different categories of bloggers' professions in the sample: Alphabetical order

S No	Category	Abbreviation	Working Definition
1	Analysts	A	Columnists and analysts of all types
2	Business	B	Bloggers who belong to business, finance, accounting, and banking
3	Consultants	C	All types of consultants like business, management, educational etc.
4	Doctors	D	General practitioners and specialists
5	Electronic Media	EM	Bloggers from TV/Radio professions
6	Engineers	E	All kinds of engineers including software engineers
7	Journalists	J	All types of journalists and newspaper staff
8	Law	L	Law practitioners and judges
9	Marketing	M	Bloggers associated with marketing sector
10	Music/Film	M/F	Bloggers with a background of professions like music, film industry, theatre, photography, and visual arts
11	Not Available	N/A	Bloggers whose professional background was not available
12	Not Clearly Mentioned	N/C	Bloggers whose professions were not clearly mentioned

Table 3.3 (Continued)

Different categories of bloggers' professions in the sample: Alphabetical order

13	Politics	P	Prominent political leaders of Pakistan
14	Private-Sector Employees	PSE	Bloggers who did not specifically mention their professions but indicated to be serving or have served in this sector.
15	Public Servants	PS	Serving and retired government employees
16	Social Work	SW	Social workers and political/human rights activists
17	Sports	Sp	Leading sports players of Pakistan
18	Students	St	Adult students at all levels of studies as well as researchers
19	Teachers	T	Bloggers from teaching professions at any level of education from school teachers to professors
20	Writers	W	Bloggers who called themselves as writers and include those who claimed authorship of books and publications

3.3.5 Corpus Size

Corpus size is a fluid concept and there is no ideal size for a corpus. In this regard, studies have determined two major factors. One relates to the representation of the type of language being investigated while the other relates to the available resources. While some (Gatto, 2014) believe that the size of a corpus depends upon its nature and purpose, others (Reppen, 2010) argue that it revolves around two factors: corpus representativeness, i.e. to represent the type of language being investigated; and, practicality, i.e. the amount of time available to gather the data. Studies also claim that rather than the sample size in a corpus, a thorough definition of the target population is the most important consideration in building corpora (Biber, 1993).

In terms of the numbers of words in a corpus, Baker (2006) provides general guidelines based on the purpose of the research with the help of quoting relevant examples. He argues that,

for non-discourse oriented studies, a million words of a variety of language are sufficient; for analysing a discourse, a corpus may be much smaller in size; and, for analysing a particular genre of language, relatively small sample of language will be sufficient.

The present study also focused on the analysis of gendered language differences in the genre of e-newspaper blogs, which required building two small-sized *specialised* gendered corpora for comparative analysis—a distinctive methodological advantage of *specialised* corpora (Koteyko, 2014). GenCorB included machine-readable texts selected according to an external criterion (mentioned in Section 3.2.4 *Corpus Representativeness* of this study). GenCorB consisted of the whole texts of all the posts of the Pakistani bloggers: MenCorB contained about 5.6 million words whereas WenCorB had around 3.3 million words. In all, 11258 posts were included in this study. Of these, 6706 posts were of 1674 men bloggers whereas the remaining 4552 posts were of 1212 women bloggers. The average length of post per blogger was 3386.80 for MenCorB and 2801.30 for WenCorB. The study included blogs posted between November 01, 2008 and August 31, 2015.

3.4 Methods: Corpus Analysis

For analysis of language differences between men and women bloggers, all the files of blog posts were clustered into one file for each blogger for both MenCorB and WenCorB. This kind of arrangement was suitable because the study focused on language differences among e-newspaper bloggers on the basis of hypothesis test across 93 LIWC2015 variables for which means and standard deviations of 1674 files of men bloggers and 1212 files of women bloggers were required to be calculated separately for each corpus.

The corpus analysis was carried out by a combination of two corpus tools and techniques: LIWC2015 and AntConc 3.4.4 (Anthony, 2016). This combination was preferred for methodological and technical reasons. Methodologically, using a combination of tools in CL research ensures *triangulation* of techniques, which involves the use of separate sources of insight for studying a particular variable for validating findings. Speaking technically, the value of any corpus depends not on its size but on its usefulness to provide requisite information to researcher, which, in turn, depends upon the capability of a corpus tool to retrieve such information (Anthony, 2013). The present study required retrieval of information from GenCorB for ATA using a suitable ATA tool at each stage. ATA has been defined as "a set of methods that automatically extract statistically manipulable information about the presence, intensity, or frequency of thematic or stylistic characteristics of textual material" (Shapiro & Markoff, 1997 as cited in Mehl & Gill, 2010).

The first stage of the analysis focused on the language features, i.e. the linguistic characteristics of GenCorB. For this purpose, the corpora were analysed across 93 dimensions of the LIWC2015 (Pennebaker, Boyd, Jordan, & Blackburn, 2015). The next stage of analysis focused on detailed information at word-choice level in both the corpora, which was carried out by AntConc 3.4.4 (Anthony, 2016). These ATA tools were selected because they are versatile in terms of covering a range of ATA needs, user-friendly, and regularly maintained and updated by groups of experts in the field. The tools were applied in the following manner to obtain the particular research information from the corpora.

3.4.1 Corpus Tools & Techniques

The first stage of the analysis focused on finding differences in language use by bloggers as shown in their personal reflections in the blog posts. This stage of the analysis was carried out by applying LIWC2015. To use this software, a full academic licence for LIWC2015 was purchased online from https://liwc.wpengine.com/. This text analysis software analyses the rate at which language is used and groups results under various categories that indicate variables related to grammatical features, psychological constructs, personal concerns, and informal language markers. LIWC2015 processes each word in a text file by matching it to an inbuilt dictionary. If a word is found to be matching with one or more of the LIWC2015 categories, the scale for the particular category is increased. Once all the words in a text file are processed in this way, percentages for all the categories are calculated by LIWC2015. LIWC has been found externally valid and internally reliable and consistent across time, topic, and text source (Pennebaker & King, 1999; Pennebaker, Boyd, Jordan, & Blackburn, 2015). It has also been found a useful tool to explore personal reflections in text genres like blogs (Friginal & Hardy, 2014).

Once the texts of MenCorB and WenCorB were analysed by LIWC2015, the results were obtained in the form of a spreadsheet containing percentages, which was generated as LIWC2015 output file. While LIWC has been used widely in ATA studies, it also has its limitations. For instance, it acquires the lexical meaning of a word but cannot acquire its contextual meaning. Also, LIWC takes a more simplistic approach and cannot show *how* a particular word is used in terms of irony, sarcasm, metaphor and multiple meanings (Oberlander & Gill, 2006; Mehl & Gill, 2010; Lin, Lin, Wen, & Chu, 2016). However, these limitations were controlled with the help of AntConc 3.4.4 (Anthony, 2016), which provided concordances, frequencies and collocations of any target expression.

3.4.2 Statistical Procedures

The quantitative part of this study involved the use of both descriptive statistics and inferential statistics. The former involved calculation of percentages of various language features found in MenCorB and WenCorB, which were worked out by the in-built function of LIWC2015. LIWC was found as a notable tool for statistical top-down approaches to text analysis (Graesser, McNamara, & Louwerse, 2011). The latter, i.e. hypothesis test, mean, standard deviation, effect size, normality of data and p-value (two-tailed) were calculated with the help of appropriate statistical measures available in the freely downloadable version of XLSTAT (Addinsoft, 2016). The tool is user-friendly and easily integrates as an add-on feature with MS Excel. The tool was downloaded from https://www.xlstat.com/en/. A detailed discussion on the application and results of CL and statistical tools and procedures is given in the following chapter, i.e. *Chapter 4: Results*.

3.5 Other Considerations

3.5.1 Recording of Additional Findings

Once the analysis was over, the results were compiled and compared to the findings of the previous studies. Alongside showing differences in the language use by men and women, this study also showed some similarities. While compiling results, apart from testing the specific hypothesis framed for this study, any additional observations or findings were also pointed out for interest of and further investigation by other researchers.

3.5.2 Ethical Considerations

Since all research on human participants involves a certain degree of ethical considerations, the research approached all the online available data with a legitimate goal to conduct academic

research as authorised under the copyright law of Pakistan. Further, this study acknowledges the online services of the e-newspapers included in this study, which make such a huge linguistic repository readily accessible for general readers and language researchers. While the researcher is indebted to and gives due credit to all the authors who may have a desire to be acknowledged for their online 'work', the researcher has tried not to disclose the bloggers' identities at any stage of the study to ensure their anonymity.

CHAPTER 4: RESULTS

4.1 Chapter Overview

This chapter presents the overall results of this study. For this purpose, the chapter has been divided into five major sections: the nature of data, measuring normality, two-tailed nonparametric measurement, Bonferroni's Correction, and Cohen's *d* respectively. The chapter comprises quantitative data for different statistical tests conducted during data analysis stage alongside presenting the rationale for choosing each type of test.

4.2 Nature of Data

The first goal of the study was to find answer to a simple question: whether or not men and women bloggers of English e-newspaper blogs in Pakistan use language differently. For this purpose, both MenCorB and WenCorB were analysed. The analysis was carried out by a corpus analysis tool, LIWC2015. LIWC is a versatile and well-tested ATA tool, which has extensively been applied in CL studies (as discussed in *Chapter 2: Literature Review* of this study).

In this analysis of primary data involving two gendered corpora, the aim of the study was to see whether the results conformed to or contradicted those of the previous studies. The study expected that the results would be consistent with those of the previous research (the *difference* approach), i.e. men would use language for *information* while women would use language for *social interaction*. However, since the early studies in the *difference* approach were based on limited size of data and because of the new developments by the *dynamic* approach, no specific predictions could be made about particular language categories. So, the hypotheses were framed in general terms and a top-down quantitative approach was followed in the analysis. The study also expected that significant differences would be found on function words as these words are

useful markers to see how individuals relate to the world around them (Newman, Groom, Handelman, & Pennebaker, 2008). Additionally, the study looked for a wide range of words showing psychological and social processes, which include words that refer to family, friends, understanding, and emotions to see which gender use them significantly more in comparison to the other.

To carry out the analysis, separate text (.txt) files were prepared for all blog posts by each male and female blogger. In this way, a total of 1,674 files of men bloggers (MenCorB) and 1,212 files of women bloggers (WenCorB) were obtained for an equal number of bloggers in the respective corpora. LIWC2015 was used to process the files of MenCorB and WenCorB alternatively. The software compares grapheme patterns of the input files from both the corpora with those incorporated in its default dictionary containing 6,549 words (Wissen & Boot, 2017). In the present study, the tool analyzed the target files across 93 variables representing various language features. These features have been grouped under 8 major dimensions, i.e. *Word Count*, *Summary Language Variables*, *Linguistic Dimensions*, *Other Grammar*, *Psychological Processes*, *Time Orientation*, *Personal Concerns*, and *Punctuation*; each of these dimensions, in turn, has its own categories (or language features) supported by sub-dictionaries. If a word in the target corpora was found in any of the LIWC2015 categories, the scale for that language category or feature was incremented. LIWC2015 yielded results in percentages, which were exported to separate MS Excel files for MenCorB and WenCorB as represented here by Figures 4.1 and 4.2. A sample of these results is given at *Appendix B* to this study.

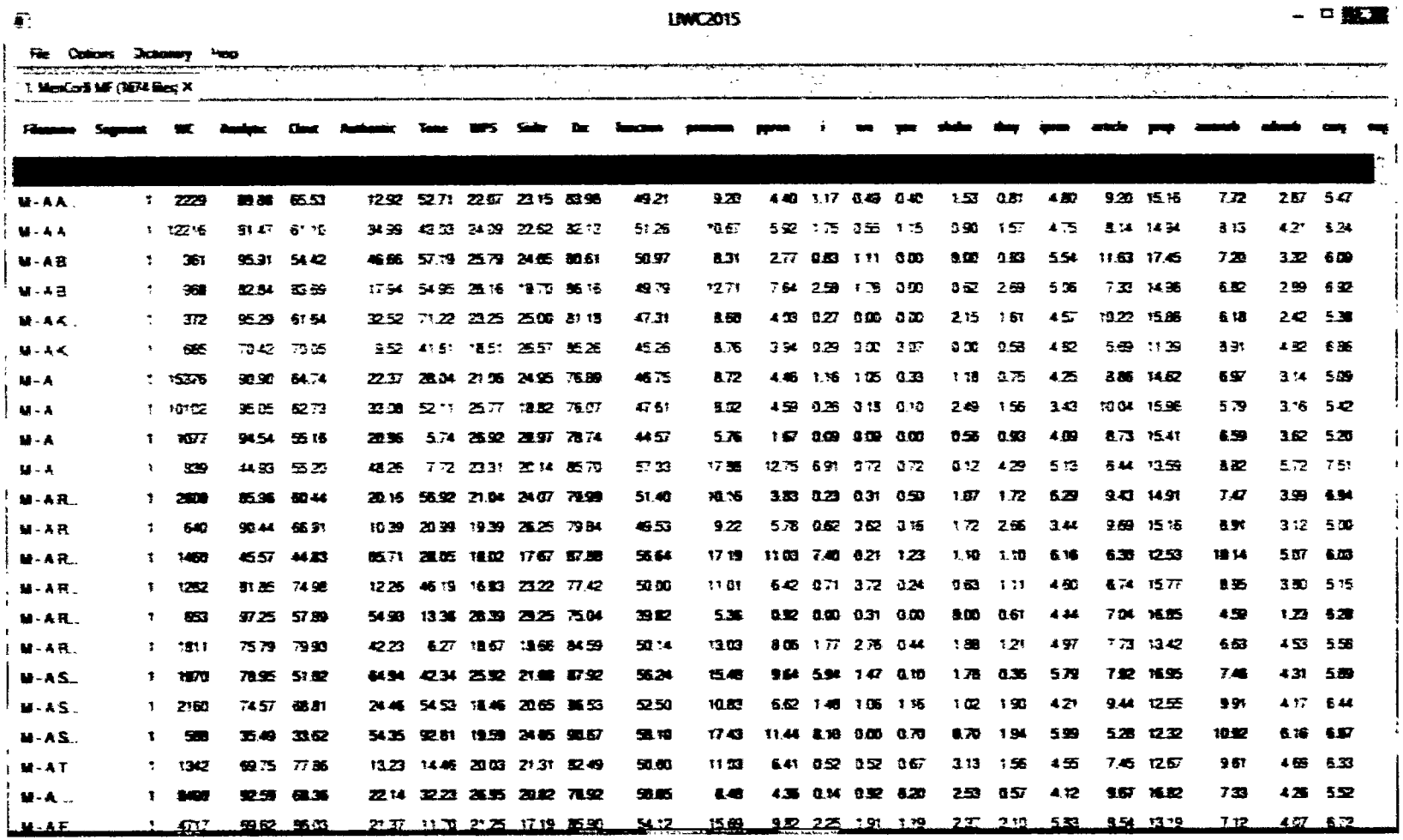
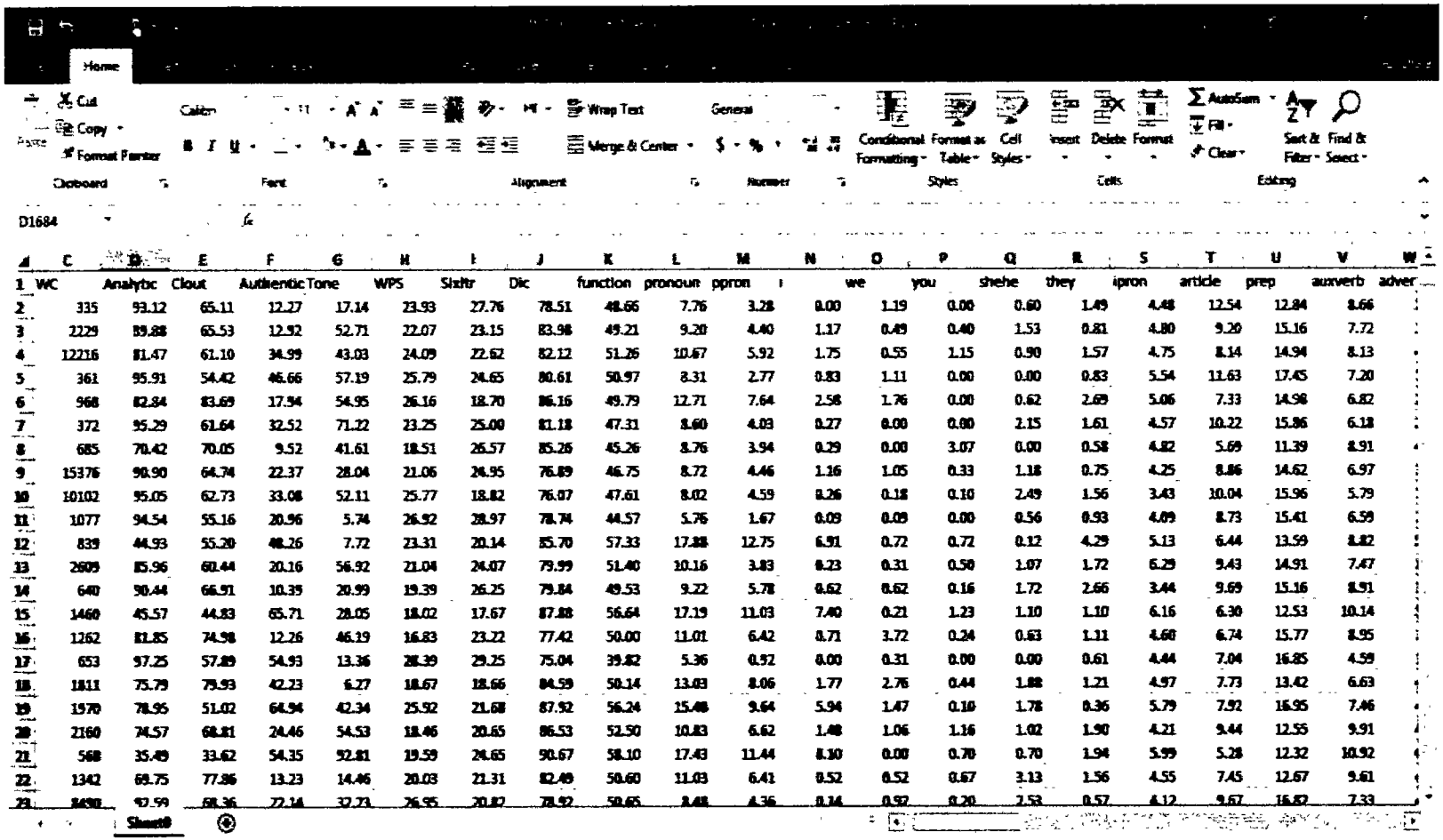


Figure 4.1. Screenshot showing sample LIWC2015 results display. The column of file names has been squeezed for this screenshot to ensure bloggers' anonymity.



	WC	Analytic	Clout	Authentic	Tone	WPS	Sixltr	Dic	function	pronoun	ppron	i	we	you	shehe	they	ipron	article	prep	auxverb	adver
2	335	93.12	65.11	12.27	17.14	23.93	27.76	78.51	48.66	7.76	3.28	0.00	1.19	0.00	0.60	1.49	4.48	12.54	12.84	8.66	
3	2229	89.88	65.53	12.92	52.71	22.07	23.15	83.98	49.21	9.20	4.40	1.17	0.49	0.40	1.53	0.81	4.80	9.20	15.16	7.72	
4	12216	81.47	61.10	34.99	43.03	24.09	22.62	82.12	51.26	10.67	5.92	1.75	0.55	1.15	0.90	1.57	4.75	8.14	14.94	8.13	
5	361	95.91	54.42	46.66	57.19	25.79	24.65	80.61	50.97	8.31	2.77	0.83	1.11	0.00	0.00	0.83	5.54	11.63	17.45	7.20	
6	968	82.84	83.69	17.94	54.95	26.16	18.70	86.16	49.79	12.71	7.64	2.58	1.76	0.00	0.62	2.69	5.06	7.33	14.98	6.82	
7	372	95.29	61.64	32.52	71.22	23.25	25.00	81.18	47.31	8.60	4.03	0.27	0.00	0.00	2.15	1.61	4.57	10.22	15.86	6.18	
8	685	70.42	70.05	9.52	41.61	18.51	26.57	85.26	45.26	8.76	3.94	0.29	0.00	3.07	0.00	0.58	4.82	5.69	11.39	8.91	
9	15376	90.90	64.74	22.37	28.04	21.06	24.95	76.89	46.75	8.72	4.46	1.16	1.05	0.33	1.18	0.75	4.25	8.86	14.62	6.97	
10	10102	95.05	62.73	33.08	52.11	25.77	18.82	76.07	47.61	8.02	4.59	0.26	0.18	0.10	2.49	1.56	3.43	10.04	15.96	5.79	
11	1077	94.54	55.16	20.96	5.74	26.92	28.97	78.74	44.57	5.76	1.67	0.09	0.09	0.00	0.56	0.93	4.09	8.73	15.41	6.59	
12	839	44.93	55.20	48.26	7.72	23.31	20.14	85.70	57.33	17.88	12.75	6.91	0.72	0.72	0.12	4.29	5.13	6.44	13.59	8.82	
13	2609	85.96	60.44	20.16	56.92	21.04	24.07	79.99	51.40	10.16	3.83	0.23	0.31	0.50	1.07	1.72	6.29	9.43	14.91	7.47	
14	640	90.44	66.91	10.35	20.99	19.39	26.25	79.84	49.53	9.22	5.78	0.62	0.62	0.16	1.72	2.66	3.44	9.69	15.16	8.91	
15	1460	45.57	44.83	65.71	28.05	18.02	17.67	87.88	56.64	17.19	11.03	7.40	0.21	1.23	1.10	1.10	6.16	6.30	12.53	10.14	
16	1262	81.85	74.98	12.26	46.19	16.83	23.22	77.42	50.00	11.01	6.42	0.71	3.72	0.24	0.63	1.11	4.60	6.74	15.77	8.95	
17	653	97.25	57.89	54.93	13.36	28.39	29.25	75.04	39.82	5.36	0.92	0.00	0.31	0.00	0.00	0.61	4.44	7.04	16.85	4.59	
18	1811	75.79	79.93	42.23	6.27	18.67	18.66	84.59	50.14	13.03	8.06	1.77	2.76	0.44	1.88	1.21	4.97	7.73	13.42	6.63	
19	1570	78.95	51.02	64.94	42.34	25.92	21.68	87.92	56.24	15.48	9.64	5.94	1.47	0.10	1.78	0.36	5.79	7.92	16.95	7.46	
20	2160	74.57	68.81	24.46	54.53	18.46	20.65	86.53	52.50	10.83	6.62	1.48	1.06	1.16	1.02	1.90	4.21	9.44	12.55	9.91	
21	568	35.49	33.62	54.35	92.81	19.59	24.65	90.67	58.10	17.43	11.44	8.10	0.00	0.70	0.70	1.94	5.99	5.28	12.32	10.92	
22	1342	68.75	77.86	13.23	14.46	20.03	21.31	82.49	50.60	11.03	6.41	0.52	0.52	0.67	3.13	1.56	4.55	7.45	12.67	9.61	
23	8490	92.59	68.36	22.14	32.73	26.95	20.87	78.92	50.65	8.48	4.36	0.14	0.92	0.20	2.53	0.57	4.12	9.67	16.82	7.33	

Figure 4.2. Screenshot of sample LIWC2015 results exported to MS Excel file. The column of file names has not been included in this screenshot to ensure bloggers' anonymity.

After initial descriptive statistics about the two samples of MenCorB and WenCorB were obtained in the form of percentages of the language features across LIW2015 variables and their means and standard deviations were calculated, an appropriate statistical measure was required for arriving at a decision on the null hypothesis and the alternate hypothesis framed for this study.

The choice of appropriate statistical test for making inferences about a sample of data depends upon the assumption a test makes about the distribution of data in the sample. Parametric tests (e.g. *t-test*) assume that the data in a sample follows normal distribution and compares *means* whereas their nonparametric counterparts (e.g. *Mann-Whitney U-test*) suit skewed values that deviate normal distribution (Neideen & Brasel, 2007; Brezina, 2018). Therefore, before selecting a test for analysis, normality of data had to be verified because by simply assuming that a data in a sample follows normal distribution without empirical evidence, there was a serious risk of arriving at wrong results should such an assumption proves to be incorrect. For this purpose, data distribution was checked empirically in both the samples of both MenCorB and WenCorB.

4.3 Measuring Normality

Normality in data can be assessed both graphically and analytically. Graphically, histograms, Probability-Probability (PP) plots and Quantile-Quantile (QQ) plots are used to see normal distribution. XLSTAT (Addinsoft, 2016) generates these plots for data to test normality. Since they are a subjective way of assessing normality, combining such plots with an appropriate analytical measure is generally recommended (Das & Imon, 2016). Several studies (Anderson & Darling, 1954; D'Agostino, 1986; Ponocny, 2001; Thode, 2002; Chaseling, Steele, Smart, & Hurst, 2012) have mentioned different goodness-of-fit analytical tests, namely Kolmogorov-

Smirnov, Shapiro-Wilk, Vasicek, Anderson-Darling, Kupier, Jarque-Bera and Cramer von Mises, which can be used to see if data in a sample is distributed normally.

Studies have also been conducted on power comparison of these tests via Monte Carlo simulation with contradictory results (Shapiro, Wilk, & Chen, 1968; Mendes & Pala, 2003; Keskin, 2006; Farrel & Stewart, 2006; Razali & Wah, 2010; Yap & Sim, 2011; Das & Imon, 2016). Taking all these considerations, this study combined Jarque-Bera (JB) tests with the graphical technique (histogram, PP plots and QQ plots) to see the normality in the sampled data.

JB test is goodness-of-fit measure of whether or not a sample data has the skewness and kurtosis matching a normal distribution. The test measured data for both samples of MenCorB and WenCorB at JB critical value 5.991 with 2 degree of freedom (df). Results of the analytic measure, i.e. JB test, for all LIWC2015 variables are shown in Table 4.1.

Generally, for JB test, the null hypothesis (H_0) suggests a normal distribution. To test the null hypothesis for normality, this study referred to the p-value. So, in the results shown in Table 4.1, p-value >0.05 means that the data are normally distributed whereas *p-value* <0.05 means the data are not normally distributed. In case of *p-value* <0.05, the study accepted the alternate hypothesis (H_1) for normality, i.e. the data did not have normal distribution. As can be seen in Table 4.1, the *p-value* for a two-sample test for all the 93 LIWC2015 variables suggest that the data differed significantly from normal distribution.

Table 4.1

Results of Jarque-Berra normality test for all LIWC2015 language dimensions

	LIWC Category	MenCorB				WenCorB			
		Mean	SD	Observed JB Value	p-value (Two-tailed)	Mean	SD	Observed JB Value	p-value (Two-tailed)
1	Word Count	3386.80	14148.05	76347109.096	< 0.0001	2801.30	5581.19	158074.711	< 0.0001
Summary Language Variables									
2	Analytical thinking	83.22	12.91	1736.561	< 0.0001	75.73	16.06	291.045	< 0.0001
3	Clout	66.24	11.79	82.952	< 0.0001	69.53	14.18	174.974	< 0.0001
4	Authentic	27.76	16.01	762.844	< 0.0001	32.27	19.54	189.915	< 0.0001
5	Emotional tone	41.93	23.73	59.269	< 0.0001	39.51	25.41	64.349	< 0.0001
6	Word/sentence	22.01	4.50	1046.801	< 0.0001	20.82	4.27	48.536	< 0.0001
7	Words > 6 letters	22.77	4.10	28.545	< 0.0001	21.25	4.22	48.035	< 0.0001
8	Dictionary words	81.39	4.51	48.867	< 0.0001	84.14	4.64	35.950	< 0.0001
Linguistic Dimensions									
9	Total function words	49.73	3.81	7.405	0.025	51.32	4.10	22.935	< 0.0001
10	Total pronouns	10.11	3.34	332.720	< 0.0001	12.03	3.74	34.362	< 0.0001
11	Personal pronouns	5.38	2.94	641.718	< 0.0001	7.04	3.37	93.988	< 0.0001
12	1st person singular	1.36	1.76	5699.527	< 0.0001	2.28	2.27	678.002	< 0.0001
13	1st person plural	1.04	0.93	1675.502	< 0.0001	1.17	1.02	1949.357	< 0.0001
14	2nd person	0.55	0.95	29246.241	< 0.0001	0.93	1.20	7829.624	< 0.0001
15	3rd pers singular	1.32	1.47	9946.032	< 0.0001	1.46	1.63	3182.881	< 0.0001
16	3rd pers plural	1.12	0.76	2787.839	< 0.0001	1.21	0.81	4944.604	< 0.0001
17	Impersonal pronouns	4.73	1.20	306.474	< 0.0001	4.98	1.16	134.691	< 0.0001
18	Articles	8.76	1.58	15.323	0.000	8.02	1.54	34.535	< 0.0001
19	Prepositions	14.71	1.47	67.109	< 0.0001	14.44	1.49	11.062	0.004

Table 4.1 (Continued)

Results of Jarque-Berra normality test for all LIWC2015 language dimensions

LIWC Category		MenCorB				WenCorB			
		Mean	SD	Observed JB Value	p-value (Two-tailed)	Mean	SD	Observed JB Value	p-value (Two-tailed)
20	Auxiliary verbs	7.51	1.44	15.199	0.001	7.70	1.45	22.159	< 0.0001
21	Common Adverbs	3.99	1.03	50.227	< 0.0001	4.36	1.11	55.409	< 0.0001
22	Conjunctions	5.90	1.04	255.072	< 0.0001	6.25	1.01	23.183	< 0.0001
23	Negations	1.32	0.55	196.286	< 0.0001	1.49	0.64	318.473	< 0.0001
Other Grammar									
24	Common verbs	12.77	2.34	22.662	< 0.0001	13.75	2.53	12.988	0.002
25	Common adjectives	4.61	0.97	86.399	< 0.0001	4.52	0.96	108.973	< 0.0001
26	Comparisons	2.50	0.70	241.414	< 0.0001	2.39	0.67	385.749	< 0.0001
27	Interrogatives	1.49	0.57	177.522	< 0.0001	1.69	0.61	149.388	< 0.0001
28	Numbers	1.85	1.21	24006.767	< 0.0001	1.54	0.98	7137.517	< 0.0001
29	Quantifiers	2.16	0.61	132.257	< 0.0001	2.21	0.65	458.692	< 0.0001
Psychological Processes									
30	Affective processes	5.15	1.35	330.906	< 0.0001	5.44	1.44	345.765	< 0.0001
31	Positive emotion	2.93	1.05	902.928	< 0.0001	2.98	1.11	497.569	< 0.0001
32	Negative emotion	2.15	1.05	2372.325	< 0.0001	2.38	1.18	508.889	< 0.0001
33	Anxiety	0.38	0.33	348313.666	< 0.0001	0.43	0.32	3226.402	< 0.0001
34	Anger	0.72	0.61	5286.864	< 0.0001	0.74	0.61	2457.138	< 0.0001
35	Sadness	0.44	0.33	4982.797	< 0.0001	0.51	0.41	12829.279	< 0.0001
36	Social processes	8.52	3.00	304.620	< 0.0001	10.40	3.59	68.309	< 0.0001
37	Family	0.32	0.57	45378.822	< 0.0001	0.70	0.90	6968.003	< 0.0001

Table 4.1 (Continued)

Results of Jarque-Berra normality test for all LIWC2015 language dimensions

	LIWC Category	MenCorB				WenCorB			
		Mean	SD	Observed JB Value	p-value (Two-tailed)	Mean	SD	Observed JB Value	p-value (Two-tailed)
38	Friends	0.18	0.21	163500.564	< 0.0001	0.24	0.25	4658.342	< 0.0001
39	Female references	0.46	0.99	129422.353	< 0.0001	1.29	1.65	4549.305	< 0.0001
40	Male references	1.36	1.43	3520.728	< 0.0001	1.21	1.34	11369.020	< 0.0001
41	Cognitive processes	10.34	2.19	145.738	< 0.0001	10.92	2.31	116.291	< 0.0001
42	Insight	1.99	0.75	607.496	< 0.0001	2.19	0.82	948.718	< 0.0001
43	Causation	1.73	0.66	566.876	< 0.0001	1.76	0.65	737.609	< 0.0001
44	Discrepancy	1.35	0.62	422.434	< 0.0001	1.41	0.60	362.122	< 0.0001
45	Tentative	2.21	0.83	2102.910	< 0.0001	2.36	0.82	123.118	< 0.0001
46	Certainty	1.47	0.52	177.727	< 0.0001	1.57	0.56	167.119	< 0.0001
47	Differentiation	2.94	0.87	206.727	< 0.0001	3.08	0.96	182.520	< 0.0001
48	Perceptual processes	1.79	0.96	4147.737	< 0.0001	2.29	1.14	1338.464	< 0.0001
49	See	0.79	0.59	56771.224	< 0.0001	0.96	0.65	11384.185	< 0.0001
50	Hear	0.51	0.50	19551.630	< 0.0001	0.62	0.51	6644.058	< 0.0001
51	Feel	0.36	0.30	25680.667	< 0.0001	0.50	0.40	34479.960	< 0.0001
52	Biological processes	1.41	1.31	46904.499	< 0.0001	2.15	1.64	3987.870	< 0.0001
53	Body	0.47	0.49	41602.029	< 0.0001	0.68	0.63	5855.253	< 0.0001
54	Health	0.59	0.81	85243.312	< 0.0001	0.82	0.86	14333.996	< 0.0001
55	Sexual	0.06	0.20	581448.960	< 0.0001	0.12	0.33	300736.938	< 0.0001
56	Ingestion	0.28	0.62	598719.433	< 0.0001	0.50	1.05	44194.433	< 0.0001

Table 4.1 (Continued)

Results of Jarque-Berra normality test for all LIWC2015 language dimensions

LIWC Category		MenCorB				WenCorB			
		Mean	SD	Observed JB Value	p-value (Two-tailed)	Mean	SD	Observed JB Value	p-value (Two-tailed)
57	Drives	8.66	2.03	181.437	< 0.0001	8.31	2.15	116.406	< 0.0001
58	Affiliation	2.33	1.20	661.316	< 0.0001	2.63	1.41	2378.570	< 0.0001
59	Achievement	1.87	1.01	3917.088	< 0.0001	1.52	0.72	1096.124	< 0.0001
60	Power	3.86	1.51	405.384	< 0.0001	3.27	1.37	413.436	< 0.0001
61	Reward	1.16	0.61	5230.678	< 0.0001	1.10	0.53	1129.197	< 0.0001
62	Risk	0.69	0.45	8691.598	< 0.0001	0.67	0.42	1124.518	< 0.0001
Time Orientation									
63	Past focus	3.68	1.73	432.278	< 0.0001	3.71	1.97	216.156	< 0.0001
64	Present focus	7.65	2.06	47.146	< 0.0001	8.39	2.15	24.850	< 0.0001
65	Future focus	1.00	0.51	607.980	< 0.0001	0.97	0.51	21242.295	< 0.0001
66	Relativity	13.47	2.25	28.712	< 0.0001	13.17	2.21	39.531	< 0.0001
67	Motion	1.69	0.62	552.258	< 0.0001	1.74	0.70	4153.704	< 0.0001
68	Space	7.33	1.50	176.982	< 0.0001	7.01	1.46	85.930	< 0.0001
69	Time	4.55	1.31	181.442	< 0.0001	4.53	1.27	250.034	< 0.0001
Personal Concerns									
70	Work	3.43	1.92	1770.554	< 0.0001	2.79	1.89	3635.276	< 0.0001
71	Leisure	1.32	1.18	4265.040	< 0.0001	1.18	0.95	2315.657	< 0.0001
72	Home	0.29	0.32	13662.679	< 0.0001	0.43	0.39	2343.683	< 0.0001
73	Money	0.85	1.08	40260.768	< 0.0001	0.65	0.71	60285.116	< 0.0001
74	Religion	0.63	0.88	20507.096	< 0.0001	0.61	0.86	26569.414	< 0.0001
75	Death	0.30	0.43	60630.959	< 0.0001	0.31	0.44	6074.054	< 0.0001
73	Money	0.85	1.08	40260.768	< 0.0001	0.65	0.71	60285.116	< 0.0001

Table 4.1 (Continued)

Results of Jarque-Berra normality test for all LIWC2015 language dimensions

	LIWC Category	MenCorB				WenCorB			
		Mean	SD	Observed JB Value	p-value (Two-tailed)	Mean	SD	Observed JB Value	p-value (Two-tailed)
74	Religion	0.63	0.88	20507.096	< 0.0001	0.61	0.86	26569.414	< 0.0001
75	Death	0.30	0.43	60630.959	< 0.0001	0.31	0.44	6074.054	< 0.0001
Informal Language									
76	Informal language	0.30	0.29	16261.468	< 0.0001	0.33	0.30	16317.236	< 0.0001
77	Swear words	0.03	0.09	643821.906	< 0.0001	0.03	0.07	80136.900	< 0.0001
78	Netspeak	0.08	0.19	155286.018	< 0.0001	0.08	0.17	394637.544	< 0.0001
79	Assent	0.08	0.13	902844.521	< 0.0001	0.09	0.12	12980.869	< 0.0001
80	Nonfluencies	0.12	0.12	3191.732	< 0.0001	0.12	0.14	8697.114	< 0.0001
81	Fillers	0.01	0.02	168396.098	< 0.0001	0.01	0.03	165049.197	< 0.0001
Punctuation									
82	Total Punctuation	13.55	2.63	679.938	< 0.0001	14.27	2.78	716.445	< 0.0001
83	Period	4.48	1.00	4591.971	< 0.0001	4.61	1.01	227.618	< 0.0001
84	Comma	4.63	1.37	4296.892	< 0.0001	4.80	1.38	13715.598	< 0.0001
85	Colons	0.22	0.29	67371.118	< 0.0001	0.22	0.28	8705.301	< 0.0001
86	Semicolons	0.18	0.22	10524.388	< 0.0001	0.21	0.22	1125.446	< 0.0001
87	Question marks	0.35	0.41	78645.366	< 0.0001	0.44	0.44	5619.354	< 0.0001
88	Exclamation marks	0.11	0.20	151706.422	< 0.0001	0.17	0.27	12251.894	< 0.0001
89	Dashes	0.84	0.60	5920.561	< 0.0001	0.73	0.52	1313.693	< 0.0001
90	Quotation marks	0.73	0.70	3632.890	< 0.0001	0.95	0.81	3396.004	< 0.0001
91	Apostrophes	1.28	0.74	1459.888	< 0.0001	1.47	0.85	2049.726	< 0.0001
92	Parentheses	0.60	0.63	17176.274	< 0.0001	0.53	0.54	2513.980	< 0.0001
93	Other punctuation	0.14	0.33	697541.431	< 0.0001	0.13	0.29	462220.617	< 0.0001

Since the analytic tests are not the only method of measuring normality in data, the study also checked *skewness* and *kurtosis* through visual inspection of histograms, PP plots and QQ plots for each of the 93 LIWC2015 variables of both MenCorB and WenCorB. These graphs also indicated that our data for all the 93 LIWC2015 variables in both the samples tended to deviate from normal distribution. For space limitation in this section, results for normality for all the 93 LIWC2015 variables through graphical measures could not be shown here. Therefore, results for only one of the LIWC2015 variables (i.e. Pronoun) are shown with the help of histograms, PP plots and QQ plots both for MenCorB and WenCorB respectively in Figures 4.3 to 4.8 below.

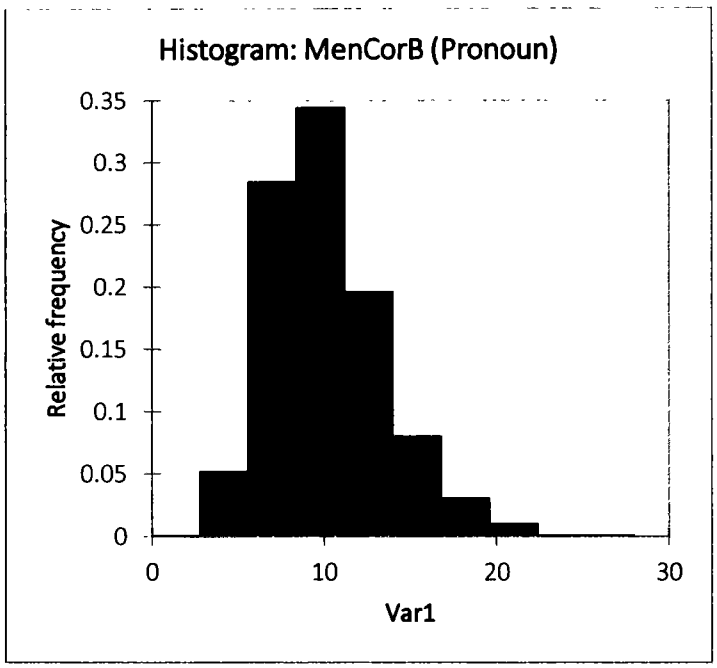


Figure 4.3. Histogram showing the non-normality of sampled data from MenCorB for LIWC2015 variable 'Pronoun'

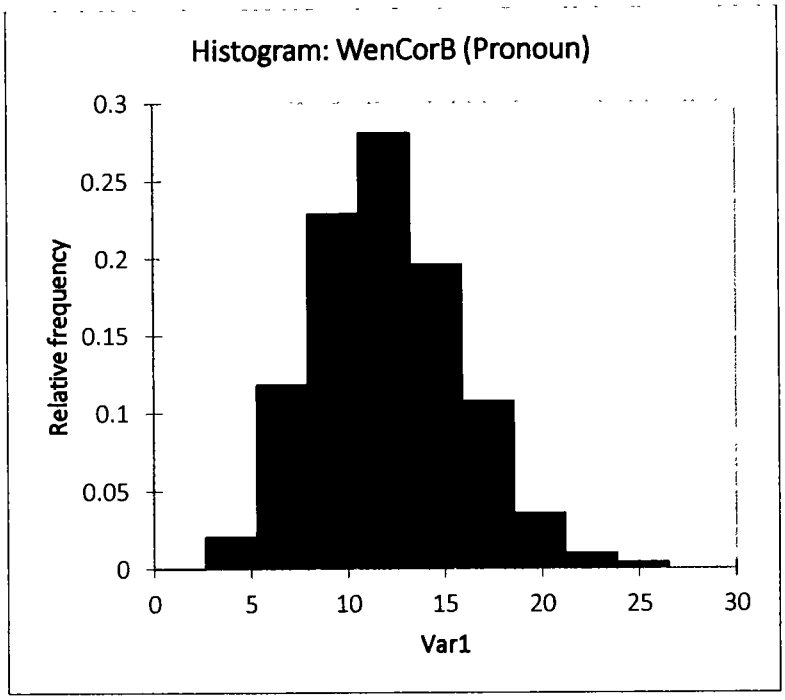


Figure 4.4: Histogram showing the non-normality of sampled data from WenCorB for LIWC2015 variable 'Pronoun'

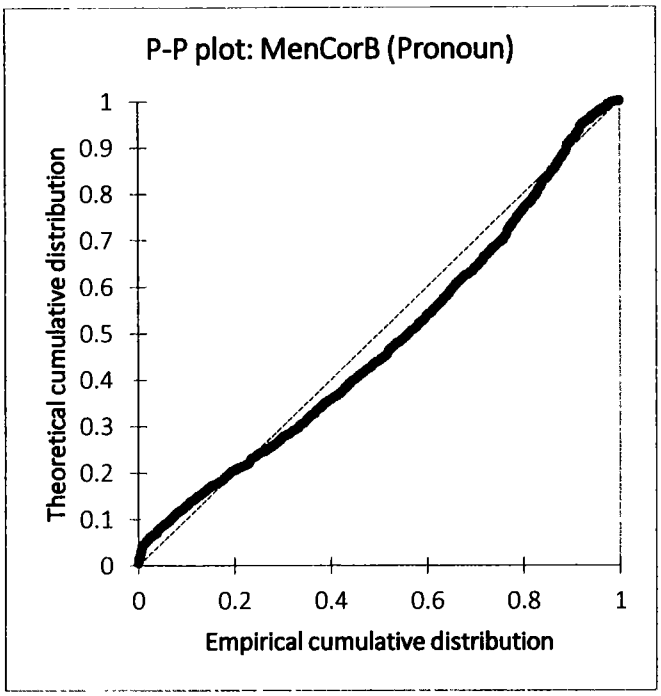


Figure 4.5: PP plot showing the non-normality of sampled data from MenCorB for LIWC2015 variable 'Pronoun'

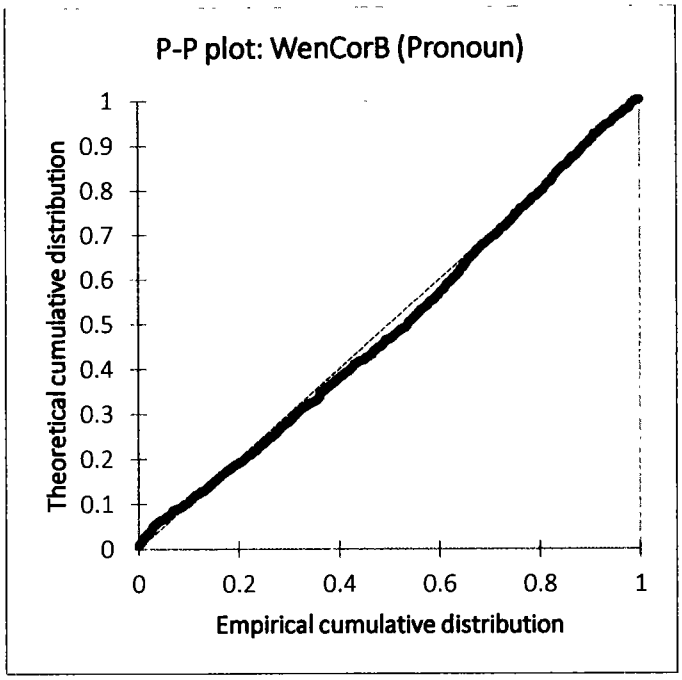


Figure 4.6: PP plot showing the non-normality of sampled data from WenCorB for the LIWC2015 variable 'Pronoun'

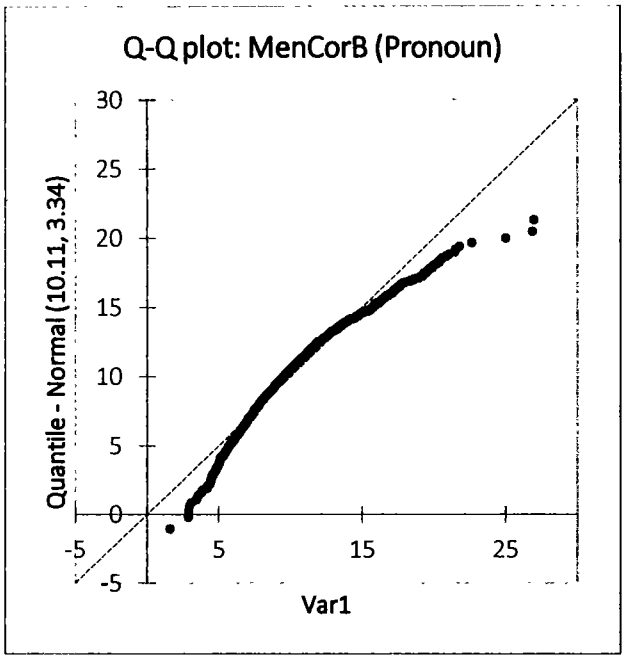


Figure 4.7: QQ plot showing the non-normality of sampled data from MenCorB for LIWC2015 variable 'Pronoun'

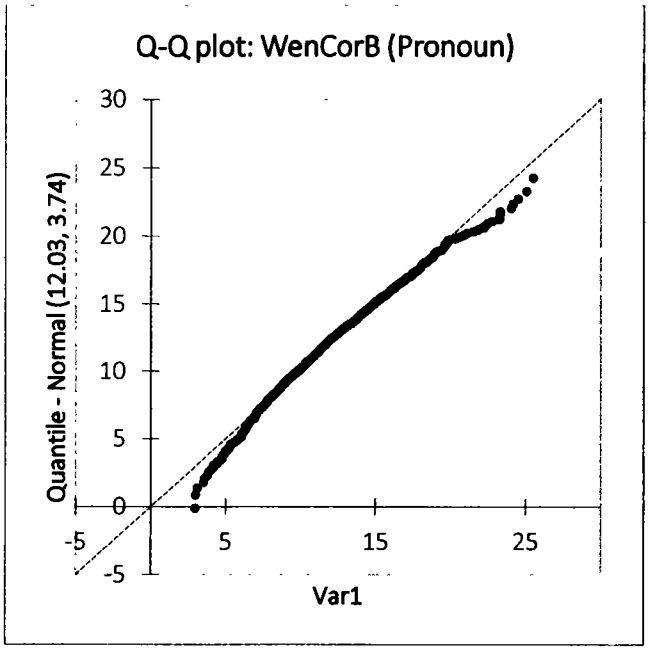


Figure 4.8: QQ plot showing the non-normality of sampled data from WenCorB for LIWC2015 variable 'Pronoun'

The results given in Table 4.1 and those depicted by histograms, PP plots and QQ plots suggested that the values obtained for LIWC2015 variables were skewed. Hence, a nonparametric measure would be appropriate for further inferential statistics for this study.

4.4 Two-Tailed Nonparametric Measurement

Since the data for the study consisted of two independent samples of MenCorB and WenCorB, a nonparametric two independent sample test *Mann-Whitney U-test*—also known as *Wilcoxon Test*, *Wilcoxon-Mann-Whitney W Test*, *Wilcoxon Rank-Sums Test*, or simply, *U Test* (Cramer & Howitt, 2004, p. 180)—was selected for testing the null hypothesis. The test is used as an alternative to its parametric equivalent, i.e. the *independent samples t-test*. While the *t-test* sets a pre-condition that the population under study must follow a particular distribution (e.g. normal distribution), Mann-Whitney U-test is a distribution-free hypothesis test, which imposes no such

condition on the population under study. Mann-Whitney U-test was preferred over t-test since samples for the present study were not normally distributed. Moreover, the Mann-Whitney U-test is suitable for a greater variety of data, has a fair efficiency rate of 0.95 in comparison to t-test and requires larger samples to test the null hypothesis (Larose, 2015). More importantly, Mann-Whitney U-test is used, as in the present study, with one *categorical independent variable* having two levels (e.g. gender: men/women) and one *ordinal dependent variable*, which can be ordered in ranks for analysis (Tavakoli, 2012), e.g. percentages of language features in the present study.

The level of risk (also called *significance*) or *alpha* (α) associated with the null hypothesis was set at 0.05. In other words, the present study set that there was a 95% chance that the observed statistical differences would be real and not due to chance. All the computational work was done with the help of XLSTAT (Addinsoft, 2016) add-in package for MS Excel and results for each of the 93 LIWC2015 variables were calculated one by one. Complete results for Mann-Whitney U-test are given at *Appendix C* to this study.

4.5 Bonferroni's Correction

Although results for each of the 93 LIWC2015 variables were obtained at the set alpha level, these could not be applied to the whole family of the LIWC2015 variables to decide about the null hypothesis. Therefore, to avoid family-wise Type I error, the *Bonferroni-Dunn* test (also called as *Bonferroni's Correction*) was conducted. The test is based on the *Bonferroni's inequality*, which states that "the probability of the occurrence of a set of events can never be greater than the sum of the individual probabilities for each event" (Sheskin, 2000, p. 541). The test is computationally identical to multiple t-tests, Fisher's LSD or linear contrasts. It is calculated by

Table 4.2 shows the results of means and standard deviations, Mann-Whitney U-test (after Bonferroni's Correction), and Cohen's effect sizes for both the corpora. The Table has been divided into five major columns. The first and the second columns show the serial numbers and nomenclature of LIWC2015 categories and sub-categories. The third column, i.e. Summary Statistics, is bifurcated to reflect means ($\overline{X}$) and standard deviations (σ) each for MenCorB and WenCorB. All means, except *word count* and *words per sentence*, are percentages of word counts in each category in MenCorB and WenCorB. The fourth column represents *p-values* obtained after U-Test while the final column shows effect sizes (Cohen's *d*) for LIWC2015 categories where language of the genders different significantly. Positive effect sizes show men's overuse of the category whereas negative effect sizes show that of women's. All categories with *p-values* (after Bonferroni's Correction) > .0005 (rounded-off) were considered not significant (*ns*).

Table 4.2

Hypotheses tests and main effects of gender on language

S No	LIWC Category	Summary Statistics				*p-value*	*d*
		MenCorB		WenCorB			
		$\overline{X}1$	σ1	$\overline{X}2$	σ2		
Word Count							
1	Word Count	3386.80	14148.05	2801.30	5581.19	0.3526	ns
Summary Language Features							
2	Analytical thinking	83.22	12.91	75.73	16.06	<0.0001	0.52
3	Clout	66.24	11.79	69.53	14.18	<0.0001	-0.25
4	Authentic	27.76	16.01	32.27	19.54	<0.0001	-0.25
5	Emotional tone	41.93	23.73	39.51	25.41	0.0019	ns
6	Word/sentence	22.01	4.50	20.82	4.27	<0.0001	0.27
7	Words > 6 letters	22.77	4.10	21.25	4.22	<0.0001	0.37
8	Dictionary words	81.39	4.51	84.14	4.64	<0.0001	-0.60
Linguistic Dimensions							
9	Total function words	49.73	3.81	51.32	4.10	<0.0001	-0.40
10	Total pronouns	10.11	3.34	12.03	3.74	<0.0001	-0.54
11	Personal pronouns	5.38	2.94	7.04	3.37	<0.0001	-0.53

Table 4.2 (Continued)

Hypotheses tests and main effects of gender on language

S No	LIWC Category	Summary Statistics				*p-value*	*d*
		MenCorB		WenCorB			
		$\overline{X}1$	σ1	$\overline{X}2$	σ2		
12	1st person singular	1.36	1.76	2.28	2.27	<0.0001	-0.46
13	1st person plural	1.04	0.93	1.17	1.02	<0.0001	-0.13
14	2nd person	0.55	0.95	0.93	1.20	<0.0001	-0.35
15	3rd pers singular	1.32	1.47	1.46	1.63	0.0414	ns
16	3rd pers plural	1.12	0.76	1.21	0.81	0.0007	ns
17	Impersonal pronouns	4.73	1.20	4.98	1.16	<0.0001	-0.21
18	Articles	8.76	1.58	8.02	1.54	<0.0001	0.48
19	Prepositions	14.71	1.47	14.44	1.49	<0.0001	0.19
20	Auxiliary verbs	7.51	1.44	7.70	1.45	<0.0001	-0.13
21	Common Adverbs	3.99	1.03	4.36	1.11	<0.0001	-0.34
22	Conjunctions	5.90	1.04	6.25	1.01	<0.0001	-0.34
23	Negations	1.32	0.55	1.49	0.64	<0.0001	-0.30
Other Grammar							
24	Common verbs	12.77	2.34	13.75	2.53	<0.0001	-0.41
25	Common adjectives	4.61	0.97	4.52	0.96	0.0129	ns
26	Comparisons	2.50	0.70	2.39	0.67	<0.0001	0.15
27	Interrogatives	1.49	0.57	1.69	0.61	<0.0001	-0.33
28	Numbers	1.85	1.21	1.54	0.98	<0.0001	0.29
29	Quantifiers	2.16	0.61	2.21	0.65	0.0313	ns
Psychological Processes							
30	**Affective processes**	5.15	1.35	5.44	1.44	<0.0001	-0.21
31	Positive emotion	2.93	1.05	2.98	1.11	0.3703	ns
32	Negative emotion	2.15	1.05	2.38	1.18	<0.0001	-0.21
33	Anxiety	0.38	0.33	0.43	0.32	<0.0001	-0.18
34	Anger	0.72	0.61	0.74	0.61	0.2321	ns
35	Sadness	0.44	0.33	0.51	0.41	<0.0001	-0.20
36	**Social processes**	8.52	3.00	10.40	3.59	<0.0001	-0.57
37	Family	0.32	0.57	0.70	0.90	<0.0001	-0.52
38	Friends	0.18	0.21	0.24	0.25	<0.0001	-0.25
39	Female references	0.46	0.99	1.29	1.65	<0.0001	-0.63
40	Male references	1.36	1.43	1.21	1.34	0.0073	ns
41	**Cognitive processes**	10.34	2.19	10.92	2.31	<0.0001	-0.26
42	Insight	1.99	0.75	2.19	0.82	<0.0001	-0.25

Table 4.2 (Continued)

Hypotheses tests and main effects of gender on language

S No	LIWC Category	Summary Statistics				*p-value*	*d*
		MenCorB		WenCorB			
		$\overline{X}1$	σ1	$\overline{X}2$	σ2		
43	Causation	1.73	0.66	1.76	0.65	0.1132	ns
44	Discrepancy	1.35	0.62	1.41	0.60	<0.0001	-0.11
45	Tentative	2.21	0.83	2.36	0.82	<0.0001	-0.18
46	Certainty	1.47	0.52	1.57	0.56	<0.0001	-0.19
47	Differentiation	2.94	0.87	3.08	0.96	<0.0001	-0.15
48	**Perceptual processes**	1.79	0.96	2.29	1.14	<0.0001	-0.48
49	See	0.79	0.59	0.96	0.65	<0.0001	-0.27
50	Hear	0.51	0.50	0.62	0.51	<0.0001	-0.23
51	Feel	0.36	0.30	0.50	0.40	<0.0001	-0.38
52	**Biological processes**	1.41	1.31	2.15	1.64	<0.0001	-0.50
53	Body	0.47	0.49	0.68	0.63	<0.0001	-0.37
54	Health	0.59	0.81	0.82	0.86	<0.0001	-0.27
55	Sexual	0.06	0.20	0.12	0.33	<0.0001	-0.23
56	Ingestion	0.28	0.62	0.50	1.05	<0.0001	-0.27
57	**Drives**	8.66	2.03	8.31	2.15	<0.0001	0.17
58	Affiliation	2.33	1.20	2.63	1.41	<0.0001	-0.24
59	Achievement	1.87	1.01	1.52	0.72	<0.0001	0.40
60	Power	3.86	1.51	3.27	1.37	<0.0001	0.41
61	Reward	1.16	0.61	1.10	0.53	0.0388	ns
62	Risk	0.69	0.45	0.67	0.42	0.5078	ns
63	Past focus	3.68	1.73	3.71	1.97	0.3434	ns
64	Present focus	7.65	2.06	8.39	2.15	<0.0001	-0.35
65	Future focus	1.00	0.51	0.97	0.51	0.0807	ns
66	**Relativity**	13.47	2.25	13.17	2.21	0.0001	0.13
67	Motion	1.69	0.62	1.74	0.70	0.1370	ns
68	Space	7.33	1.50	7.01	1.46	<0.0001	0.22
69	Time	4.55	1.31	4.53	1.27	0.9318	ns
Personal Concerns							
70	Work	3.43	1.92	2.79	1.89	<0.0001	0.34
71	Leisure	1.32	1.18	1.18	0.95	0.2522	ns
72	Home	0.29	0.32	0.43	0.39	<0.0001	-0.39
73	Money	0.85	1.08	0.65	0.71	<0.0001	0.23
74	Religion	0.63	0.88	0.61	0.86	0.3031	ns
75	Death	0.30	0.43	0.31	0.44	0.9633	ns

Table 4.2 (Continued)

Hypotheses tests and main effects of gender on language

S No	LIWC Category	Summary Statistics				*p-value*	*d*
		MenCorB		WenCorB			
		$\overline{X}1$	σ1	$\overline{X}2$	σ2		
Informal Language							
76	Informal language	0.30	0.29	0.33	0.30	0.0016	ns
77	Swear words	0.03	0.09	0.03	0.07	0.8902	ns
78	Netspeak	0.08	0.19	0.08	0.17	0.3534	ns
79	Assent	0.08	0.13	0.09	0.12	0.0001	-0.13
80	Nonfluencies	0.12	0.12	0.12	0.14	0.2738	ns
81	Fillers	0.01	0.02	0.01	0.03	0.0276	ns
Punctuation							
82	Total Punctuation	13.55	2.63	14.27	2.78	<0.0001	-0.27
83	Period	4.48	1.00	4.61	1.01	0.0007	ns
84	Comma	4.63	1.37	4.80	1.38	0.0002	-0.12
85	Colons	0.22	0.29	0.22	0.28	0.7834	ns
86	Semicolons	0.18	0.22	0.21	0.22	<0.0001	-0.12
87	Question marks	0.35	0.41	0.44	0.44	<0.0001	-0.23
88	Exclamation marks	0.11	0.20	0.17	0.27	<0.0001	-0.26
89	Dashes	0.84	0.60	0.73	0.52	<0.0001	0.19
90	Quotation marks	0.73	0.70	0.95	0.81	<0.0001	-0.30
91	Apostrophes	1.28	0.74	1.47	0.85	<0.0001	-0.24
92	Parentheses	0.60	0.63	0.53	0.54	0.0035	ns
93	Other punctuation	0.14	0.33	0.13	0.29	0.5493	ns

CHAPTER 5: DISCUSSIONS

5.1 Chapter Overview

The chapter presents a mix of quantitative figures interpreted through qualitative discussion. Besides comparing findings of the present study to those of the previous research, this chapter shows results for further probe of the data through collocations, concordances and frequency counts of particular expressions to explore gender similarities and differences in word choices and themes/sub-themes. In this manner, results for each of the 93 language variables analysed have been elaborated and discussion has been made for all major categories of LIWC2015, i.e. *word count*, *summary language features*, *linguistic dimensions*, *other grammar*, *psychological processes*, *informal language*, and *total punctuation*. Some of these categories have many subordinate categories, which have been listed and discussed under their respective superordinate categories.

5.2 Word Count

The issue of verbosity has been discussed at length by previous research on language and gender. For instance, Swacker (1976) studied differences in language use between genders in question-answers during academic conferences. She concluded that women tended to contribute less in asking questions (27.4%), asked twice as shorter questions as men, and tended not to comment or ask additional questions after a speaker's response. The findings led Swacker (1976) to conclude that women were shier in comparison to men in speaking in public. Later, Eakins and Eakins (1978) examined tape recordings of university faculty meetings. This study supported Swacker's (1976) findings except that it found men to be speaking more and for longer durations than women did.

Spender (1980) had suggested that intuitively it was believed that women should be seen and not heard. She explained that it was because women were perceived to talk more. A study by Sadker and Sadker (1985) supported such findings, which after showing video recordings of a classroom discussion, asked teachers who talked more, boys or girls. The teachers responded that they believed that girls talked more; in reality, boys had talked thrice more than girls.

Gender differences in verbosity or length of speech were also studied at different floor types. For instance, Edelsky (1981) studied these differences at two types of floors: singly developed where one speaker speaks at a time, and, collaboratively developed, which is open to all simultaneously. The study concluded that men talked more at the former whereas women talked more at the latter forum. These findings were supported by numerous subsequent studies (Falk, 1980; Chafe, 1995; Coates, 1996; Coates & Jordan, 1997).

A growing body of research has also studied language differences in genders in CMC settings where gender of writer is relatively anonymous and it was believed that such anonymity might lead to break traditional gender binaries (e.g. Rodino, 1997; Danet, 1998). Contrary to this belief, a number of studies in CMC found gender differences like those found in speech by the previous research. These studies concluded that women were less verbose than men in online public forums (Herring & Paolillo, 2006). Previous studies in CMC have also focused attention on the issue of the length of the blog posts. However, these studies have not reached any conclusion because of the nature of the blog, the topic involved, and the gender of the writers. For instance, Page (2011) found that men wrote longer travel blogs posts while women wrote longer personal blogs. Similarly, Coates (2003) found that men wrote longer when the topic was focused or stereotypically masculine. These, studies, however, agreed that gender does make a difference in participation on the blogosphere.

In the light of contrasting previous research findings, this study could not make any particular expectations that men and women bloggers of e-newspaper of Pakistan would differ in terms of *word count*. The Mann-Whitney U-test showed that no statistically significant difference was found in word count for both the genders(U= 1034982.5; U (standardized) = 0.930; Expected Value = 1014444; Variance (U) = 488116326.214; p-value = 0.3526). So, providing statistical evidence, the present study reports that in e-newspaper blogs, men and women produce equally longer or shorter texts.

5.3 Summary Language Features

This dimension of LIWC2015 categories summarises the overall language features found in a text. In other words, these eight variables reflect the general trends in the text. Results for these summary features will be discussed one by one with the help of relevant previous research.

5.3.1 Analytical Thinking

As argued by Gentry (1989), the concepts of *equality* and *difference* are not antithetical. So, while genders may be equal, they may be different still. Though physical differences in men and women are quite obvious and can be seen, studies have also found that both genders differ psychologically, i.e. in their thinking patterns. For example, *Cognitive-Experiental Self Theory* (CEST) states that reasoning in human beings conforms to a parallel but dual thinking pattern (Epstein, 1994, 2003). This theory argues that human beings make sense of the world through two thinking systems: *rational* and *experiental*. The former demands a high level of cognitive resource and uses logic, evidence, symbols, numbers and words that are culturally transmitted. This thinking system is considered to be slow, deliberate and analytic. The latter system, in contrast, is innate

and adaptive, which enables human beings to learn from experience (Kirkpatrick & Epstein, 1992). Previous research has frequently reported significant differences in gender with men preferring *rational* style of thinking and women preferring *experiental* reasoning (Sladek, Bond, & Phillips, 2010).

Speaking in the terms of linguistics, the words that we use also indicate *what* we are paying attention to or *how* we are thinking (Tausczik & Pennebaker, 2010). Especially, the use of function words has been associated with *how* people think. Within the broader category of function words, the tendency to use more articles and prepositions has been associated with "categorical language", which has been explained as a language style that characterizes heightened abstract thinking and cognitive complexity. The former characteristic of the categorical language is represented by articles and the latter by prepositions (Pennebaker, Chung, Frazee, Lavergne, & Beaver, 2014).

The Mann-Whitney U-test results of the present study also show that men and women bloggers differ significantly in their thought patterns (U = 1324954.500; U (standardized = 14.054; Expected value = 1014444.000; Variance (U) = 488116474.127; p-value = 0.0001). As shown in Table 4.2, the language of men bloggers indicate that they think more analytically than women bloggers. Summary results for *analytical thinking* are supported by results for *function words, articles* and *prepositions* given in Table 5. Men bloggers tend to use high percentage of articles and prepositions in their blogs. So, the findings of the present study support the previous research, which has shown that men are more analytical than women.

5.3.2 Clout

This summary feature of LIWC2015 suggests the overall tendency in a language sample as an indicator of the status or rank of individuals in a social group. Previous research shows that

language can serve as a marker of social status. For instance, Lakoff (1975) argued that powerless speech uses more tag questions, hedges and intensifiers in comparison to powerful speech. Similarly, O'Barr (1982) compared trial transcripts and concluded that individuals with low status (e.g. defendants and witnesses) used greater number of hedges, intensifiers, polite forms, and hesitation forms.

More recent research has found function words to be indicators of individual's psychological states. Studies show that as style markers in a particular language, function words tend to reflect personality, emotional states, and social relationship (Chung & Pennebaker, 2007). Lexically fewer than 500 in count, English function words make up approximately 55% of the words that are used in speaking or writing (Rochon, Saffran, Berndt, & Schwartz, 2000). Within function words, the use of pronoun indicates whether attention is on others—second person (*you*) and third person (*he*, *she*, *they*) or on ourselves (*I*) or ourselves embedded within a social relationship (*we*) (Zimmerman, Wolf, Bock, Peham, & Benecke, 2013). For instance, previous research shows that, in comparison to low-status individuals, high-status individuals used "we" more frequently (Sexton & Helmreich, 2000; Cassell, Huffaker, Tversky, & Ferriman, 2006). Higher use of "we" reflects the fact that high-status individuals are more collectively oriented or other-oriented. Studies also suggest that there is a relationship between higher use of the second person pronoun (*you*) and focus on the other person and show that higher status individuals focus their attention outward, towards the person they are speaking with. Research also indicates that those who attain status have been found to take care of group success and remain more focused on others, i.e. they are more generous, collectively focused, cooperative, fair, create rapport, forge connections, resolve conflicts, and provide more help (Thibaut & Kelley, 1959; Ridgeway & Diekema, 1989; Boehm, 1999; Flynn, Reagans, Amanatullah, & Ames, 2006; Blader & Chen,

2012). In language and gender research, studies have also shown that women prefer more prestigious forms than men (Labov, 1966, 1972a; Trudgill, 1972) because women are more status conscious (Trudgill, 1972).

Clout dimension of LIWC2015 deals with the overall status-orientation of individuals in a language sample. The tool compiles results for this summary dimension on the basis of the use of personal pronouns (Pennebaker, Boyd, Jordan, & Blackburn, 2015). The Mann-Whitney U-test results for the present study given in Table 4.2 show that women tended to indicate a high status on this language dimensions (U = 834571.500; U (standardized) = -8.141; Expected value = 1014444.000; Variance (U) = 488116484.361; p-value = 0.0001). As can be seen in Table 4.2, women bloggers have used more first person plural and second person pronouns, which indicate a greater attention towards collectivity and others, which have been shown to be indicators of high status individuals. Moreover, high status has also been associated with greater sense of rapport, conflict resolution, forging connections, and cooperation, which have been found to be characteristics of women by the previous studies. Therefore, in these respects the findings of the present study support those of the previous research.

5.3.3 Authentic/Dictionary Words

These two LIWC2015 variables are treated together in this discussion. Authenticity of a text according to all LIWC versions relates to whether a text shows genuine authorship or shows features that lead analysts to conclude that the text is fake or tells lies. LIWC analyses this feature on the basis of use of words that relate to *pronouns*, *psychological processes* (like affective or emotional processes, negative emotions, causation, insight etc) and *relativity* (space, motion tenses etc) (Newman, Pennebaker, Berry, & Richards, 2003). Therefore, the summary language feature

authentic will be stronger or weaker for any text with relation to analysis of style features that form authenticity for LIWC. Since the present study did not aim to verify the authenticity or otherwise of MenCorB and WenCorB, no further discussion is made on *authentic* variable of LIWC2015.

The next variable within the group of summary language features of LIWC2105 is *dictionary words.* LIWC2015 analyses sample of texts with the help of an inbuilt dictionary of words, which is made up of sub-dictionaries for each sub-category of LIWC2015 variables that tap words, which belong to a particular language feature. Except words that show punctuation, the entire lexis of LIWC2015 is termed as *dictionary words*. Therefore, this summary variable shows the overall percentage of words in GenCorB that matched lexis in LIWC2015 categories. As percentages for this variable in Table 4.2 show, a reasonable size, i.e. more than 80% of expressions in both the corpora matched LIWC2015 dictionaries. This percentage is in line with LIWC's general trends as indicated by Newman, Groom, Handelman, and Pennebaker (2008), who report that the tool recognises about 80% of words in a target text sample while the remaining 20% are proper nouns and very low freuency words. Based on this, it could be claimed that texts in GenCorB were adequately recognizable by LIWC and results for the present study are fairly valid. Moreover, it was found theat women had a higher percentage of dictionary words (84.14%) in comparison to men (81.39%). This trend is consistent with Lester and Leenarsb (2016) who had also reported a higher percentage of dictionary words in women's text.

5.3.4 Emotional Tone

A persistent idea that has transpired over the years is that women are more emotionally expressive than men (Fischer, 1993). There are a number of studies that have looked into differences between men and women on the basis of their emotional tone. Folk wisdom claims

that, in comparison to men, women are more emotional, or, at least, are more emotionally expressive (Kring & Gordon 1998), which also has support in much of the academic research (e.g. Ashmore & Boca, 1979; Johnson & Shulman 1988; Brody & Hall 2000). Contrary to this, the present study found that statistically there was no significant difference in the overall emotional tone of men and women bloggers U = 1082952.500; U (standardized) = 3.101; Expected value = 1014444.000; Variance (U) = 488113270.495; p-value = 0.0019).

One possible explanation for this summary feature, i.e. *emotional tone,* may be that LIWC2015 has based its analysis for this variable on calculations for a number of linguistic sub-features that relate to use of words expressing *positive emotions*, *cognitive processing*, *social orientation*, and *psychological distancing* in a target text. This summary variable, i.e. *emotional tone* calculates the overall tendency of a text for percentages of words used for these sub-categories. For a detailed discussion on how LIWC factorizes these features to *emotional tone* of a text, Cohn, Mehl, and Pennebaker (2004) is a useful resource.

This explanation of the present study is also supported by previous research, which shows that certain emotions are expressed more by one gender than the other. Emotions of happiness, sadness and fear are believed to be more characteristic of women, whereas men are believed to be more characteristically angry (Kelly & -Comeaux, 1999). Similarly, as argued by Schmitt (2015) while women may be termed as emotionally expressive, men can also be called so: the difference lies in the type of emotion expressed by either gender. Therefore, such differences in emotions can neither be considered as totally absent, nor can they make the two genders unrelated to one another, i.e. the old Mars vs Venus claptrap.

Therefore, for this summary language dimension, this study supports the view of the previous research that there are certain types of emotions that men and women tend to express

differently. Results for tendency of genders in expressing particular emotions are discussed in relevant sections in this study.

5.3.5 Words Per Sentence

Previous research on relationship between gender and sentence length yielded contradictory results in a variety of communication settings. For instance, Mulac and Lundell (1994) studied description of landscape photographs by 40 students at university-level education. Their results were consistent with what they had found in an earlier research (Mulac, Lundell, & Bradac, 1986). Both the studies found that women used longer sentences in comparison to men. In another study on gender differences on giving subordinate criticism, Mulac, Seibold, and Farris (2000) found that men used more words overall while women used longer sentences. In a more recent study, Mulac, Giles, Bradac, and Palomares (2013) described men's language as showing higher dynamism, while women's reflecting higher on socio-intellectual status. They made this conclusion on the basis that women used longer sentences, hedges, intensive adverbs, and dependent clauses. In contrast to these findings, some studies also found different results. For example, analysing eight-minute interactions on problem-solving, which involved 108 university students (54 women and 54 men), Mulac (1989) found that men spoke in longer utterances. Later, Mulac, Studley, and Blau (1990) also studied fourth-grade students' essays and found that boys used longer sentences than girls.

This contradiction in results pertaining to gender and sentence length is also found in studies on blogs. In research on blogosphere, there are a number of empirical studies devoted to blog mining or gender specific text analysis. These studies have tried to explore how gender affects writing style and topic using texts from blogosphere. Sentence length has been one of the features

that the past research on blogs has taken into consideration. However, the results of the studies are contradictory. While some have found it as an indicator of gender in authorship analysis of blogs, others have reported no significant difference between genders on the basis of average sentence length. For instance, studies by Argamon, Koppel, Pennebaker, and Schler (2007), and Goswami, Sarkar, and Rustagi (2009) chose sentence length as one of the factors to distinguish between bloggers' age and gender. According to findings of these studies, the average sentence length did not correlate with age or gender. One possible reason for such findings may be the fact highlighted by Peersman, Daelemans, and Vaerenbergh (2011) that these studies worked with text fragments containing a minimum of 250 words. Contrary to the findings of these reports, a study by Zhang and Zhang (2010) found significant differences in average sentence length of blog posts on the basis of gender. The study found that, on the average, male bloggers used longer sentences containing more words (19.0883) than female bloggers (17.8864).

The above discussion shows that previous research has given contradictory results in various settings as regards the effect of gender on sentence length. The present study, however, found significant differences in sentence length on the basis of gender (U = 1175314.500; U (standardized) = 7.281; Expected value = 1014444.000; Variance (U) = 488115973.369; p-value = 0.0001). The results show a tendency in men bloggers for using longer sentences than women bloggers and support the previous research with similar findings.

5.3.6 Words > 6 Letters

The study of word length in linguistics has about 167 years long history. It was Augustus de Morgan, an English logician and mathematician, who brought forth the idea of judging an

individual's style and authorship by word length, specifically the number of letters per word in a letter to a friend on August 18, 1851 (Grzybek, 2007).

Research in the field of gender differences in language use has also paid attention to the length of words. For example, examining a corpus of one million words composed of US publications from 1961, Kučera and Francis (1967) found that women used longer words than men. Such results were, however, not supported by any of the studies included in this discussion. In contrast, later research found that men tended to use longer words in comparison to women. For instance, conducting factor analysis on various text registers Biber, Conrad, and Reppen (1998) found that, in comparison to women, men used longer words and nouns as well as used more complex language with informational and uninvolved style. Similarly, Koppel, Argamon, and Shimoni (2002) used BNC texts to predict author's gender on word use. Claiming 80% accuracy in determining gender by the technique, the study concluded that men's writing was characterised with use of longer words. In the same vein, Newman, Groom, Handelman, & Pennebaker (2008) examined a database of over 14,000 text files of texts comprising both written text (93%) and speech transcription (7%) for a period of 22 years (1980-2002) from 70 studies and 22 laboratories of the United States, New Zealand, and the United Kingdom. The research findings were the same as those of the previous research that men used longer words compared to women. Similarly, in a corpus-based analysis involving different genres and topics, Mikros (2013) also found that men used words that were 4, 8, 9, and 10 lettered more in comparison to women whereas women used 2 and 3 lettered words more in comparison to men. The study explained that this tendency could be due to men using more content words (generally larger in size) and women using more functions words, which are generally 2, 3 letter words. In online setting, Corney (2003) examined an email corpus to predict the gender of the email sender. Besides other stylometric features, he included

word length as a language feature to study. Reporting a prediction accuracy of 70.1%, he found word length to be a significant gender indicator. Similarly, predicting gender from blog posts, Zhang and Zhang (2010) found that men used longer words than women.

The Mann-Whitney U-test conducted for the present study indicated significant gender differences in the use of words longer than six letters. The present study also found that men bloggers used longer than six letter words more in comparison to women bloggers (U = 1239280; U (standardized) = 10.177; Expected value = 1014444.000; Variance (U) = 488116027.100; p-value = 0.0001). The results for the present study are, therefore, consistent with those of the previous research. In the light of the discussion, it can be concluded that word length may be considered as a significant gender indicator. The possible interpretation for this trend among men may be that they use more content words (e.g. nouns) and women use more function words (e.g. pronouns) as explained by Mikros (2013).

5.4 Linguistics Dimensions

A total of 15 different language variables are included in this category of LIWC2105. These features predominantly relate to the use of function words. Separate discussion on each of these variables follows in the succeeding sub-sections.

5.4.1 Function Words

As argued by Corver and Riemsdijk (2001), since the earliest studies in linguistics, categorization of lexical items has been a source of development in description and theory of the science of language. A central classification in syntactic categories of language has been categorising words into *content words* and *function words*. While content words (also called as lexical or substantive categories or open class words) have a semantic role in a text, function words

(also called closed class words) have non-conceptual meaning in a text. Content words include nouns, verbs, adjectives and adverbs and their function is to convey the content of communication. Function or style words, on the other hand, are pronouns, prepositions, articles, auxiliary verbs, conjunctions, negations and quantifiers. Their role is to connect, organize and relate content words to one another.

In the areas of relationship between gender and the use of the whole range of function words, there are relatively fewer studies conducted previously. Previous research shows that the overuse of function words overall is a characteristic of women's style. For instance, Biber, Conrad, and Reppen (1998) found that women used function words category more in comparison to men. Hypothesising the overuse of function words by women, Newman, Groom, Handelman, and Pennebaker (2008) also found that women's style tended to have more function words than men's style.

The characteristic of overall function words as a dominant feature of female style has also been used by studies for authorship identification in online interactions. For example, Schler, Koppel, Argmamon and Pennebaker (2005) built and analysed a corpus of blogs comprising 300 million words. While the study reported significant differences in the use of content and function words by male and female bloggers, it found that the latter used more function words category in comparison to the former. A more recent research by Werlen (2015) on tweets tested LIWC as a tool for authorship identification. The study confirmed previous findings for LIWC as a useful tool for determining an author's gender on the basis of function words.

The findings of the present study are also consistent with those of the previous research. The results of Mann-Whitney U-test for this study given in Table 4.2 show that women bloggers tended to use more function words in comparison to men bloggers (U = 773400.500; U

5.4.3 Personal Pronouns

Exploring different grammatical features as indicators of language differences in men and women, some of the earliest studies in different settings (Gleser, Gottschalk, & Watkins, 1959; Mulac, Lundell, & Bradac, 1986) found the use of personal pronouns to be higher in women than men. In CL research, such an interaction between gender and the use of personal pronoun was observed by Koppel, Argamon and Shimoni (2002), and Argamon, Koppel, Fine and Shimoni (2003) in the study of BNC corpus. Their studies also found that women favoured the use of personal pronouns more than men. Likewise, the use of personal pronouns was also studied as a gender indicator in weblogs (Herring & Paolillo, 2006), which was consistent with the findings of the previous research.

The findings of the present study are also consistent with those of the previous research. The Mann-Whitney U-test found significant gender differences in the use of personal pronouns. This study found that women bloggers tended to use personal pronouns more than men bloggers (U = 696183; U (standardized) = -14.405; Expected value = 1014444.000; Variance (U) = 488115654.882; p-value = 0.0001).

5.4.4 First Person Singular

Gender and the use of first person singular pronoun has been investigated previously by a number of studies. The findings of these studies are contradictory: while some of them (Brownlow, Rosamond, & Parker, 2003; Newman, Groom, Handelman, & Pennebaker, 2008; Andersson, 2012) found the use of first person singular to be a dominant feature of female writing, others (Mulac, Giles, Bradac, & Palomares, 2013; Ahmad & Mehmood, 2015) reported an overuse of first person singular in texts produced by men.

The Mann-Whitney U-test conducted for the present study also found significant gender differences in the use of the first person singular pronoun. The results of the present study are consistent with those that find that women used more first person singular pronoun (U = 728651.500; U (standardized) = -12.941; Expected value = 1014444.000; Variance (U) = 487688598.120; p-value = 0.0001).

Research on blog shows that personalization is indicated by the use of first person singular pronoun in blogs posts. The use of this pronoun reflects the perspective that bloggers take to their post. The use of the first person pronoun puts the writer in the centre stage. This pronoun not only reflects the inner state and actions of the bloggers but also shows that they prefer to convey information (feelings, ideas, reactions) with a direct reference to themselves while engaging the reader in the discourse (Friginal & Hardy, 2014).

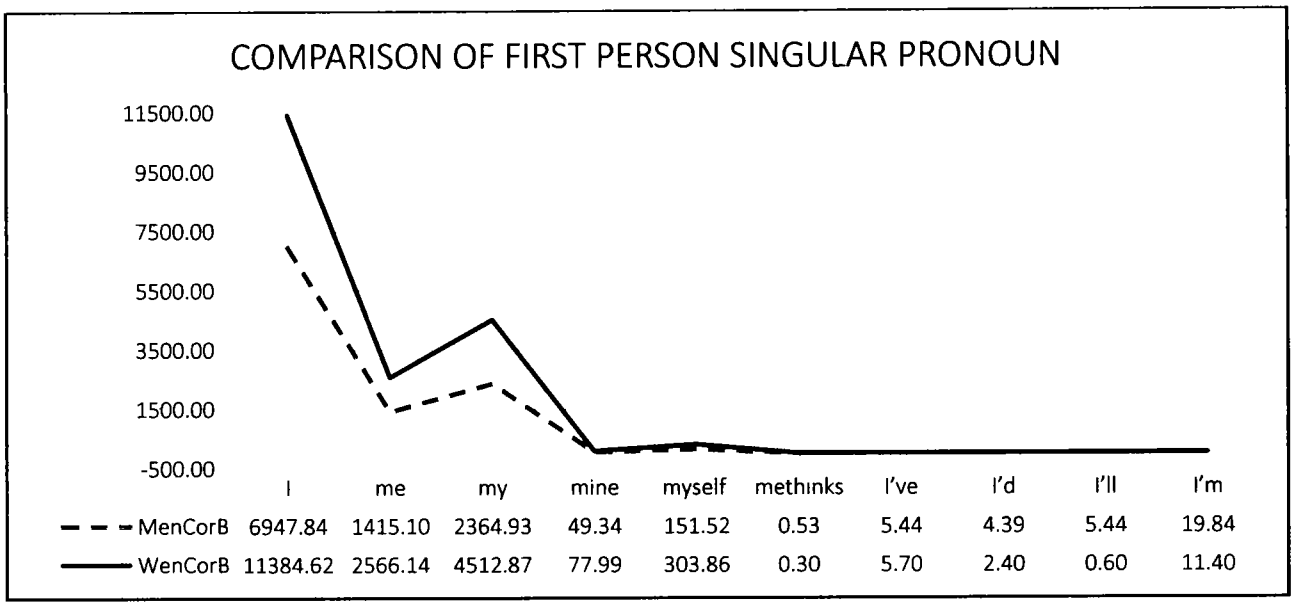


Figure 5.1. Comparison of first person singular pronoun (frequencies normalised to one million)

Once difference for the use of first person pronoun was found, a further probe was made to see how this pronoun was used differently in the two corpora. LIWC2015 captured *I, I'd, I'll, I'm, I've, me, methinks, mine, my,* and *myself* from the corpora texts to be representative tokens of

the first person singular pronoun. Figure 5.1 presents a comparative view of the normalized frequencies for these tokens in the two corpora.

Initial impressions from a high frequency of first person pronoun in Figure 5.1 show that women bloggers tend to be more self-focused in comparison to men bloggers. However, it was also observed from these results that, in comparison to men, women tended to use full forms of auxiliaries with the first person pronoun (subjective case of pronoun + auxiliary, e.g. *I had*). In contrast, men were found to use contracted forms of auxiliaries (e.g. *I'd, I'll, I'm*) and *methinks* more than women, except for *I've* structure. The overuse of contracted forms by men found by the present study supports the results of Baron (2004). Since previous research has also shown that full form subject-verb (subjective case + verb) is more emphatic as compared to contracted form(s) (Yaefer-Dror, Hall-Lew, & Deckert, 2002; Bresnan & Spencer, 2013), it could be concluded that women's writing tends to indicate comparatively more emphatic style in writing blog posts. The use of full forms by women may also be related to their liking for a formal and high-status language style, which has been reported as a feature of women's language by previous research.

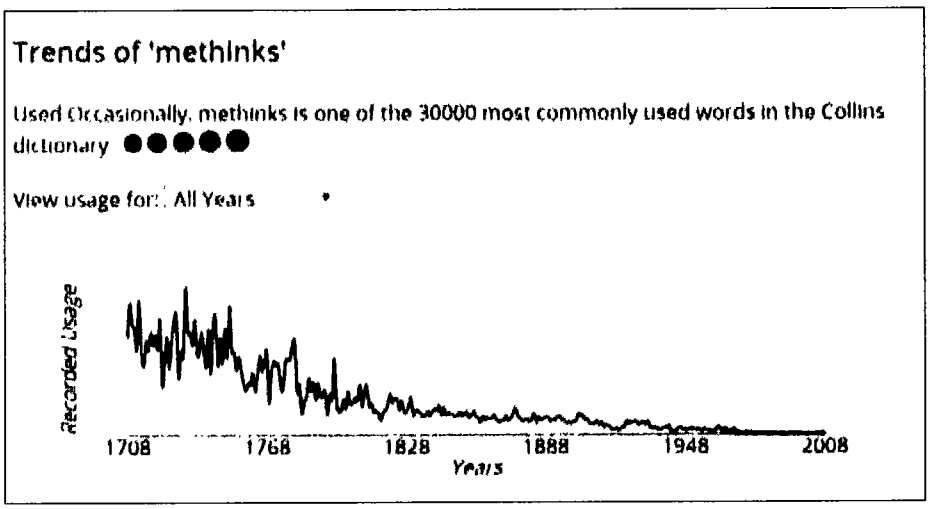


Figure 5.2. Screenshot of "Methinks" frequency over the years: Source Collins online dictionary (*https://www.collinsdictionary.com/dictionary/english/methinks*)

Diachronic research by Palander-Collin (1998), which compared the use of 'I think' and 'Methinks', reported that the latter expression was fairly frequent in the Late Middle and the Early Modern English, with a gradual decline in its frequency from 15th to 16th centuries. The research found that the expression played the role of a sociolinguistic and pragmatic marker and was used by speakers with relatively higher social status. The online version of Collins dictionary also graphically traces its use from 1708 to 2008, which has been reproduced in Figure 5.2.

The present study found negligible use of *methinks* in both the corpora. As shown in the Concordances 5.1 and 5.2 'Methinks' was used thrice by men and only once by women in the entire GenCorB.

if it were possible? It's about integrity	***methinks.***	But then again **** sahab once famously
his history and constant zoom-ins, though	***methinks***	a reported $200 million price tag is a ta
You taught me first to beg, and now	***methinks***	you teach me how a beggar should be

Concordance 5.1 'Methinks': MenCorB

statuette for Best Actress In A Supporting Role?	***Methinks***	it's **** year for both. 5) Will

Concordance 5.2 'Methinks': WenCorB

'Methinks' appears to be a register variation of 'I think' and a colloquial or speech-based device, not necessarily nonstandard, that has evidential meaning to show one's knowledge (Palander-Collin, 1999). Evidential 'think' can express *doubt* or *certainty* (Biber & Finegan , 1989) or, in Holmes' (1990) words, it can be used for *deliberative* and *tentative* purposes.

While specific studies on the use of 'methinks', particularly focusing on gender differences, are extremely rare, discussion in this section draws on insights gained from previous research on 'I think'. The analysis of Concordances 5.1 and 5.2, shows that men used 'methinks' to express more certainty. In particular, the last line in Concordance 5.1 appears to reflect that the writer is in no doubt about the intention of the addressee and uses 'methinks' as a softener to convey an affective meaning of 'I am certain'. In contrast, the only women user of 'methinks' in

the entire WenCorB seems to use the expression to signal her uncertainty or tentativeness about the possible prize winner of the contest.

Since Lakoff's (1973, 1975) earlier claims, uncertainty or tentativeness have been considered as characteristics of women's language. The findings of the present research for 'methinks' also tend to show that men and women use it for different pragmatic purposes, i.e. to show certainty and uncertainty respectively. While research on gender differences in the use of 'methinks' is almost scarce, the results of the present study could be a significant contribution to the field.

After studying these, a much closer look was taken through both the corpora for words expressing family relations that occur immediately after the first person singular possessive case (*my*). For this purpose, a range of collocates was obtained at a minimum raw frequency occurrence of 10 in each corpus. Subsequently, frequencies for the results were normalized, which have been given in Figure 5.3.

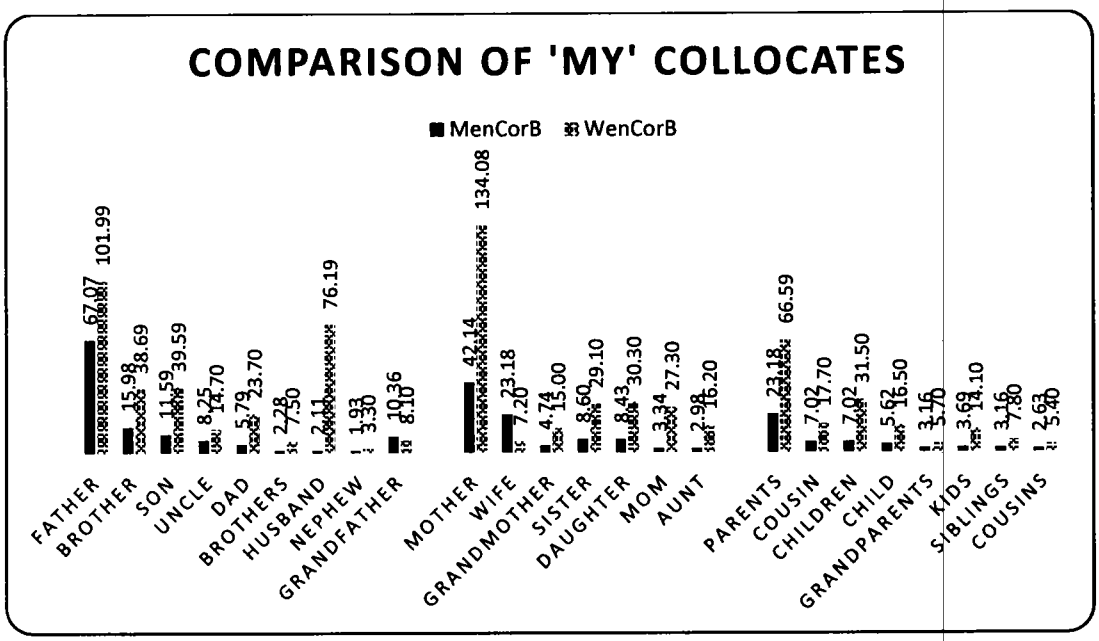


Figure 5.3. Noun referents (masculine, feminine, neuter) to family members collocating with 'My' in GenCorB (frequencies normalised to one million)

The overall picture showed that, except for the words 'my grandfather', women bloggers mention more words that relate to other family relations in their blog posts in comparison to men bloggers. These results also contrasted with those found by Schwartz, et al. (2013) who analysed a corpus of 700 million words, phrases and topic instances from volunteer Facebook users and found that men preferred to precede the words 'wife' and 'girlfriend' with first person singular possessive case (*my*) more than women did for 'boyfriend' or husband'. In the present results, the expression 'my wife' (normalised frequency = 23.18) was used with lesser frequency in MenCorB than 'my husband' (normalised frequency = 76.19) used in WenCorB. It is pertinent to add here that the words 'my husband' in men bloggers' and 'my wife' in women bloggers' texts occur where they quote some other female and male person.

Another significant observation in these results was the overuse of the words 'my grandfather' in MenCorB (normalised frequency = 10.36) in comparison to WenCorB (normalised frequency = 8.10). This highlighted the fact that men bloggers make more references to the male head of a family beyond their parental level in lineage. This tendency was further corroborated by use of words 'my ancestors' (normalised frequency = 2.45) and 'my elders' (normalised frequency = 1.75), which occurred more in MenCorB in comparison to WenCorB where the two expressions occurred with negligible frequency. One possible explanation for this tendency may be that family histories are traced after male ancestors and seen through male eyes (Ancestry, n.d.). Especially, in patriarchal societies like that of Pakistan even clans and tribes are named after common male ancestors (Metlo, 2012). Therefore, these results suggested that, in comparison to women bloggers, men bloggers tend to use more expressions that show stronger bond with ancestry and family lineage.

5.4.5 First Person Plural

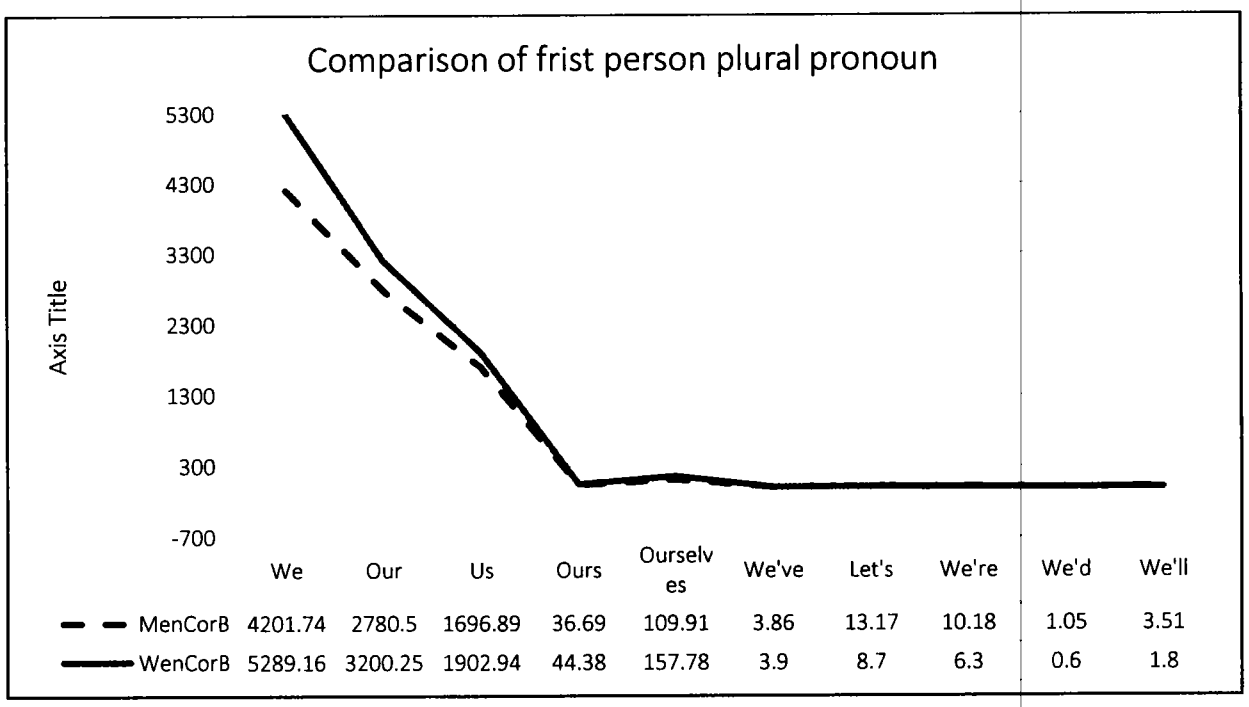
In research on gender and language difference, though previous studies (Newman, Groom, Handelman, & Pennebaker, 2008) have found overall pronoun overuse as an indicator of women's language, investigation on the use of first person plural pronoun has brought forth contrastive results. For example, Biber (1988) argued that the use of first person pronoun is the feature of 'involved' language and that the language of women display 'involvement'. However, the study of Argamon, Koppel, Fine, and Shimoni (2003) found that women used first person plural more in non-fiction texts whereas men used it more in fiction texts. Later, controlling topic and all other variables, Schmid (2015) studied the HCRC Map Task data and analysed a handful of frequencies but did not find any significant differences between men and women in the use of first person pronoun.

In online setting, Herring and Paolillo (2006) gathered texts from randomly selected blog posts and balanced the sample for author and genre (diary/filter type blog). They found that women bloggers used more first person plural pronoun in all its cases as they had hypothesised. Recently, Abdurahman (2017) randomly selected 1000 tweets each for men and women from a publicly available corpus Roverto Twitter N-Gram Corpus (RTC) of 75 million Twitter posts. The study looked for gender differences in the use of personal pronouns only. While it found that women used more pronouns overall, it observed that men tended to use more first person plural (*we* and *us*) cases.

The results of the present study support research, which found that women tend to use more first person plural pronoun in comparison to men. As reflected by the results in Table 4.2 and the calculation obtained from Mann-Whitney U-test, this study (U = 929380.500; U (standardized) = -3.851; Expected value = 1014444.000; Variance (U) = 488007733.527; p-value = 0.0001) found

that women bloggers used more first person plural pronoun in the sampled corpora. Thus, this language feature corresponds to the 'involved' characteristic of women's language as interpreted by Biber (1988).

The difference in the use of first person plural pronoun was further probed by studying the normalised frequencies of all its cases and their contracted forms captured by LIWC2015. Figure 5.4 reflects the tendencies observed. The results were consistent with those found by the present study for the first person singular pronoun cases and their contracted forms. It was found that except in one case, i.e. *we've*, women tended to overuse full forms of auxiliaries with the first person plural pronoun and men tended to overuse their contracted forms.



	We	Our	Us	Ours	Ourselves	We've	Let's	We're	We'd	We'll
MenCorB	4201.74	2780.5	1696.89	36.69	109.91	3.86	13.17	10.18	1.05	3.51
WenCorB	5289.16	3200.25	1902.94	44.38	157.78	3.9	8.7	6.3	0.6	1.8

Figure 5.4. First person plural pronoun (frequencies normalised to one million)

The greater use of plural pronouns reflects a tendency of classes. It was interesting to see that both men and women wrote texts, which showed that they talked on behalf of some institution of collectivity. To explore this, as earlier, a systematic random sample of 100 concordances each for first person plural subjective case (*we*) was studied for both the corpora. An analysis of these

concordances revealed that both men and women talked on behalf of two major and similar collective classes.

The first of the collective class was the 'nation'; both men and women bloggers expressed their feelings and attitudes on behalf of the nation as shown in Concordances 5.3 and 5.4 from MenCorB and WenCorB respectively.

company branding (any company that sponsors that signal).	*We*	can then offer any consumer brand a chance
understand. It is time to apply our minds and decide where	*we*	stand. Because wherever you stand, your camp needs y
South Asian pack when it comes to charitable giving.	*We*	contribute around Rs150 billion nearly 1% of the nation's
people earlier, and they will continue to kill until	*we*	come forward and break the vicious circle of ignorance.
beliefs. However, it is vital to our progression that	*we*	are aware of our right to question. Just as we left
media and the myth-believing nature of Pakistanis,	*we*	grow up blaming the Central Ruet-i-Hilal Committee (the
start anew. On the contrary, I'm suggesting that	*we*	make an attempt to progress on the foundation already
identify unanimously that we are not perfect and also that,	*we*	are not the greatest nation in this world. Let's discuss

Concordance 5.3 Spokesperson of 'nation': MenCorB

prescribed by the World Health Organisation. While	*we*	show our commitment to combating drug abuse,
commercials which aren't even linked together and	*we*	shouldn't take it seriously. But have you realised
so hard for us to believe in Malala's magnificence if	*we*	were a nation of people who stood up when it felt
becoming an even larger nuisance in the future? Will	*we*	allow past crimes of brutal assault against innocent
these children remain where they are, is because if	*we*	do try removing them off the streets we will have
sheep when they hardly know how to count.	*We*	give them a toy when they need our touch. Let's
ousness that Pakistan is going down, you tell them that	*we*	Pakistanis are resilient enough to rebound!
the talent we have in our country but until and unless	*we*	start getting our act together, we are never

Concordance 5.4 Spokesperson of 'nation': WenCorB

The second major group on behalf of which both men and women bloggers appeared to speak was the 'ordinary citizen' as exemplified in Concordances 5.5 and 5.6.

ew details to the general public. The end product, as	*we*	see it, has been blanked out at several places so that national
homes are working for the rehabilitation of the elderly. But	*we*	know that we cannot leave our elderly in isolation owing
family isn't pressing charges against the school then who are	*we*	to object and went on with their lives without as much
mutually exclusive? 3) Valentine's Day encourages obscenity	*We*	live in a world where we have to hide to make love,
on the basis of where they live and on their heritage.	*We*	all condemn that sort of activity but yet we practice it
that we would be veracious in answering this question. Do	*we*	really act honestly in our daily life matters? Do we try
if they can buy stuff online with few clicks. But, I guess	*we*	still have to wait more. Where Pakistan-America Relations

Concordance 5.5 Spokesperson of 'ordinary citizen': MenCorB

us achieve our goals in life are left unacknowledged.	*We*	witnessed the same situation with the coach of our national girls
, then whose fault do you think that is? Where should	*we*	put the blame? On the government I suppose! Uff that corrupt
a question on Facebook about lineage and why	*we*	should have to follow our forefathers' example. There were
mix of bittersweet events. On one hand where	*we*	hear Indo-Pak politicians finger-wagging at each other from,
praised them for all the attacks that they have stopped?	*We*	don't even know how many have lost their limbs or even lives
Zia Regime have suddenly sprung up. The schools	*we*	loved so much have become unrecognisable. Many have been
. usy we are, eating is a basic activity of everyday life.	*We*	need healthy food to function efficiently in all aspects of life so

Concordance 5.6 Spokesperson of 'ordinary citizen': WenCorB

Besides these gender similarities in representing collective classes, men and women bloggers also tended to represent different groups. As shown by Concordance 5.7, men were found to represent the national 'cricket team'.

strategies are changed. For example, I would understand if	*we*	play only seven batsmen to accommodate a third seamer.
Sarfaraz as the third opener. According to Waqar, since	*we*	did not have a third opener, we took in Nasir Jamshed.
outdoing every Pakistani team before us since 2005. Then,	*we*	came to the break rounds. We won the octofinals against
dropping Junaid we not only did not pick our best seamer,	*we*	over-burdened Gull and as a result Pakistan s seam attack
early knock-out for India in the Champions Trophy, which	*we*	could hasten when we play them in the group stage, will
nsorship appeal means hockey is given a back-seat, even if	*we*	are a winning team. Moreover, all the glamour and sca
top ranked Pakistani batsman in both ODI and T20 formats.	*We*	are wasting him at number 6 and number 7, he should bat

Concordance 5.7 Spokesperson of 'cricket team': MenCorB

However, none of the sampled concordances in WenCorB revealed women bloggers talking on behalf of the national cricket team. In Pakistani context, cricket is a popular sport form men and men's national cricket team has been playing since as early as 1952. It was only recently, i.e. in 1997, that a women's cricket team of Pakistan also made its debut. The tendency in men's blog posts shown in Concordance 5.7 is an indicator of men's liking of sport in contrast to women, which has been shown by previous research.

In contrast, women tended to speak on behalf of the same-sex group as reflected in Concordances 5.8 from WenCorB.

surrender to their manipulations regarding marriage.	*We*	can bring about a change with a little extra effort by s
I'm rambling. Just bear with me. Come on,	*we*	're girls, we should confide in each other. I really like s
er, poverty, disease and exploitation – evils that	*we*	all aspire to see an end to. Not just that, if these young girls

Concordance 5.8 Spokesperson of 'same-sex group': WenCorB

To sum up the discussion, the quantitative analysis pointed out the overuse of first person plural pronoun (*we*), its full form variants and one contracted form (*we've*) by women bloggers whereas the overuse of its remaining contracted forms (*let's*, *we're*, *we'd*, *we'll*) by men bloggers. The qualitative analysis found some similarities in how the two genders invoke common identities. The common identities were 'nation' and 'ordinary citizen' for both the genders. The genders, however, also tended to identify themselves with particular groups. In this case, men were found to talk on behalf of the national 'cricket team' whereas women appeared to represent the 'same-sex group'. These observations support somewhat similar findings by Ndambuki and Janks (2010) who concluded that women constructed their identity through the members of the gender they belonged to.

5.4.6 Second Person Pronoun

Second person pronoun is used for the person addressed to. Unlike that of other personal pronouns, the analysis of second person pronouns poses a unique problem to researchers. Since its subjective case (*you*) can be used for both singular or plural addressee, it is hard to identify the grammatical number of the person(s) spoken to or addressed to.

In language and gender research, previous studies have produced contradictory results related to second person pronoun. For instance, a study on sport reporters by Kuo (2003) found that male reporters used second person pronoun more than female reporters. Besides, the study discovered gender differences in the pragmatic use of the second person pronoun by TV sport

reporters in Taiwanese. In the same vein, another study on the language of call centres by Friginal (2009) observed that male operators used more second person pronoun than their female counterparts. Similarly, the findings of another study on editors' letters in gender-directed magazines by Andersson (2012) were also consistent with these studies.

In contrast, selecting text features for their Gender Genie, a corpus analysis tool to automatically detect author's gender, Argamon, Koppel, Fine, and Shimoni (2003) found second person pronoun to be a feature of female authored text. Their findings, however, were later not supported by research on weblogs. Testing the hypothesised overuse of second person pronoun by women in the light of Argamon, Koppel, Fine and Shimoni's (2003) earlier findings, Herring and Paolillo (2006) found that, in comparison to male bloggers, female bloggers tended to disfavour the use of second person pronoun in blogs.

The findings of the present study were consistent with those of Argamon, Koppel, Fine and Shimoni (2003). The Mann-Whitney U-test of the present study found that women bloggers used more second person pronoun in comparison to men bloggers (U = 748314; U (standardized) = -12.103; Expected value = 1014444.000; Variance (U) = 483534234.096; p-value (Two-tailed) = 0.0001).

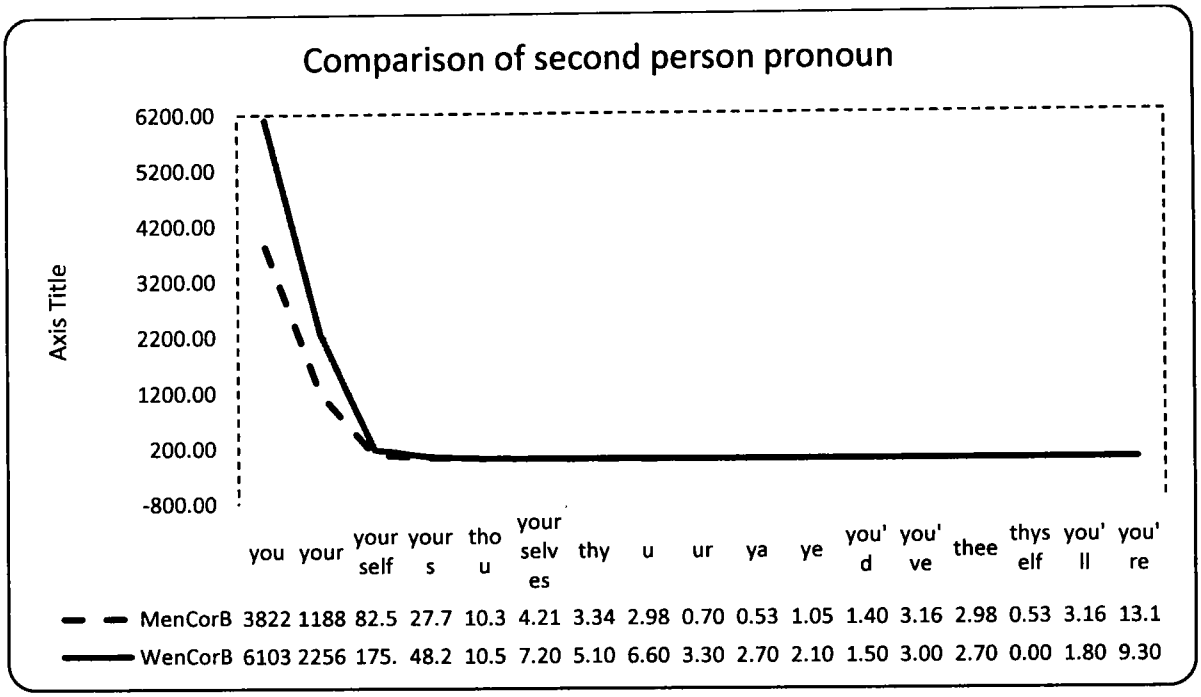


Figure 5.5. Second person pronoun variants (frequencies normalized to one million)

While the overall use of second person pronoun was favoured by women, a finer analysis of its variants for the present study revealed four categories (full forms, archaic forms, forms used in internet language, and contracted forms) for which different interesting observations were recorded. As shown in Figure 5.5, it was found the women used more second person pronoun in three variants: full forms (*you*, *your*, *yourself*, and *yourselves*), archaic forms (*thou*, *thy*) and forms used in internet language (*u*, *ur*, *ya*, *ye*). The only contracted form used more by women was *you'd*. Men, on the other hand, were found to overuse archaic forms (*thee* and *thyself*) and all the remaining contracted forms (*you've*, *you'll*, *you're*). While the archaic forms used indicate a reference to quotations from scriptures and old literary works, the forms used in internet language show the features of *netspeak* or *chatspeak*, which is used in CMC communication including blogs. Thus, these quantitative results suggest that archaic forms of second person pronoun were used by both men and women; however, women bloggers scored higher in comparison to men bloggers on second person pronoun variants that are used in *netspeak* or *chatspeak*.

In addition, a qualitative analysis of the use of second person pronoun subjective case (*you*) was carried out to see how men and women bloggers used it in the two corpora. For this purpose, a systematic random sample of 100 concordance lines each generated for *you* in both the corpora was studied. The analysis revealed that overall, the use of *you* in both the corpora can be divided into two categories: the use of *you* to mean an 'addressee'; and, its use in the sense of 'everybody'. Both of these uses have been theorised by Sacks (1992), and Malone (1997). These two types of uses in each of the corpus are exemplified by the Concordances 5.9 to 5.12 as given below.

real Muslim actually goes to heaven." "Really? Who told	***you***	that. The aalim?" "No, actually my grandmother, God bless her.
don't bring it on too much, be nice to us and let us beat	***you***	. Please? Pretty please? Aao Parhao – A teacher can use all the
, Sholay, a Bugs Life) are inspired by this movie, and if	***you***	watch it yourself, you will find many other movies will come to
travel. 'They never ask anyone to show the money. But	***you***	[are] the only Pakistani passport in the entire flight. They [will]
. is even harder to be rejected and know all the while that	***you***	might have made it if your father had proof-read your forms. Yet,
nobody told me? Well, you're a TV news anchor, aren't	***you***	? Yes, that I am, with great ratings! Breaking News! Pakistani ....
being able to prepare for the impending final exams. "Can	***you***	please bring him to my home so I can give him lessons there?"
what a powerful position would mean, I regret to inform	***you***	that being a brigadier – which isn't a bad position at all – or b
ight key? It is not working." "Oh, sorry, I forgot to brief	***you***	. You have to push the door a bit and then press the key to the
ur nation's aspirations the much-needed injection of hope.	***You***	are both winners in my books. And 10,000 adoring fans on
tart its long march in Islamabad. Yes, ****. Yes, Baajee.	***You***	can call me ****. What is the situation like there? Pretty,
what you sowed and you will enjoy all those things that	***you***	worked hard to steal. So go on vacations and drive around in

Concordance 5.9 *you* in the sense of the 'addressee': MenCorB

ry smile. "You don't have to make a decision right now,	***you***	can tell me tomorrow." **** stopped eating, and his
for treatment. Hopefully, they will survive. I am telling	***you***	all of this because contrary to what you may believe, no
ay after the match when he randomly stated, "No, I think	***you***	are wrong. There is no way they couldn't see." Even now
a woman with mehndi dyed hair and large spectacles, "	***You***	can come sit in the counselling room." The counselling r
y times. School is for rich people. You are not rich, do	***you***	understand?" "Yes." "Give me a glass of water." "Yes."
ry about us anymore. We are in good hands. We love	***you***	, father. Always will. Your beloved son, ****. No
s it going? Just wanted to...uh. Oh for heaven's sake,	***you***	know why I'm calling." Cough. "Dude, you know that it
ly a problem with her hearing. Her talent would amaze	***you***	but her story will, no doubt, dismay you. Her husband a
I expected something more and I didn't find it." Could	***you***	live there? The traveller can't dream of it. "Roads, you h
s in it. When she lost her first tooth, she asked me, "Do	***you***	think I look weird?" I said NO! Why? She said, "I don't

Concordance 5.10 you in the sense of the 'addressee': WenCorB

have a field day with that one. The formula is so simple: all	***you***	have to do is send out a mob of camera men and reporters t
f there is one thing you must learn, it is that even if fate deals	***you***	a good hand, it does not mean you give up your power to t
un-decided or not being ready for marriage? I guess most of	***you***	decent folks will say that she should not tell him that he is

tch. You know that feeling in the pit of your stomach when	***you***	see **** batting ever since the semi-final? How it
ert: contrary to the popular perception or any insinuation that	***you***	may have that its plot is somehow based around Karachi, c
u took your date to or where you found one. That was where	***you***	hung out with your friends. There was no cyber space, no s
g the number of cigarettes could be an interim measure while	***you***	are preparing to quit. Once you have reached the decided d
procedure. How can you ensure free and fair elections when	***you***	don't identify the culprits of rigged elections and punish t
things on Facebook just to get 'Likes'? In fact, how often do	***you***	refresh your page after posting something? I know for sure
self of all values – there is no space for these in this industry.	***You***	cannot excel in the corporate sector with only performance
f a bullet train passing by downtown of a metropolis at night.	***You***	only see the well-lit boulevards and tall skyscrapers whi

Concordance 5.11 *you* in the sense of 'everybody': MenCorB

that the only department you know most about is HR – but	***you***	are nowhere close to working near that department, let alone
dies. Your best childhood teachers are the ones who nurture	***you***	, and the best current teachers are the ones who give you
that the moment one turns around to look back at the things	***you***	have loved with such passion, the essence of which you have
living and create an art out of it. But you can't dish out what	***you***	can't receive in return. Would you put your parents in such
gic ones. If you rely only on questions and information fed to	***you***	by people sitting in the newsroom, you are setting yourself
two if you trust your gut. A casual friendship shouldn't make	***you***	uncomfortable – and that should always be your red flag.
ll that is lost, the images from the tragedy bring back all that	***you***	keep in that closed drawer in the corner of your mind – all
dwindling and your eyes are droopy due to sleep deprivation,	***you***	still need to keep on marching. This also means that the
ners with your family. They are a bunch of people you know	***you***	can always rely on, rant and share your happy hours with.
refrigerated in an airtight container for three to four days. If	***you***	wish to refrigerate the toffee sauce, one can do so for seven
if they want to get out of the house for a job or something,	***you***	won't be blamed for stopping them and curtailing their
ll rates, hundreds of free smses and low activation charges,	***you***	are bound to get a text message on your mobile phone about
absolutely love. 3. That feeling when you start your car and	***you***	hear a favourite song of yours playing on the radio. Its perf

Concordance 5.12 *you* in the sense of 'everybody': WenCorB

In the general sense of 'everybody', another interesting observation was made in both the corpora. In this aspect, men and women bloggers did not use *you* for 'everybody' in the sense of 'any person from any gender'. Rather they used it in the sense of 'every man' or 'every woman', i.e. their sense of 'everybody' was gender-directed. In this sense, both men and women bloggers restricted the sense of 'everybody' to the members of their own gender group as shown in Concordances 5.13 and 5.14 below. This suggests that men and women bloggers' language reflect some kind of gender polarity.

cinema in Karachi, Cineplex is restricted to families. So	***you***	can't be a guy and watch a movie in a cinema. No, this

Concordance 5.13 *you* as gender-directed 'everybody': MenCorB

girls are taught that being skinny and having perfect pearly white skin gets	***you***	love and acceptance. A media-frenzied world
ked what you have studied so far, or what your career plans are.	***You***	may also be asked if you are "aware" of all household chores.
to stick with your chosen husband and in-laws. If they decide to pound	***you***	to unconsciousness, that's just your problem. Even when h
Even though I might not be better than you, I am not any less than	***you***	either just because I cover my face. 7. Don't you feel hot in
e 'every girl is beautiful' guy: This guy's bio and picture will tell you all	***you***	need to know about him. He thinks you're beautiful. He woul
picture of you, I was with you. People thought it hypocritical, seeing that	***you***	are an actress who has time and again displayed skin as a
now. How could I not understand you? I'm sorry I never understood you.	***You***	would do all our chores and always wonder how we would

Concordance 5.14 *you* as gender-directed 'everybody': WenCorB

In a nutshell, this part of the analysis found some quantitative differences in the use of second person pronoun variants. The results showed that women tended to overuse full forms of all the subjective, objective and possessive cases of the second person pronoun, two of the archaic variants and all the variants of this pronoun that are used in internet language. In contrast, men tended to overuse two archaic forms and the contracted variants of the second person pronoun. The qualitative analysis also provided certain interesting observations pertaining to language and gender research as it revealed that the pronoun *you* is not only used for the addressee and in the sense of 'everybody' but also in the gender-directed sense of 'every man' and 'every women' by men and women respectively.

5.4.7 Third Person Singular Pronoun (*he/she*)

Previous research conducted on studying gender differences in the use of third person singular pronoun (*he* and *she*) has yielded contradictory results. For instance, exploring a large subset of the BNC across different genres of fiction and non-fiction texts, Argamon, Koppel, Fine, and Shimoni (2003) reported that females make far greater use of the personal third person singular pronoun (*he/she*) in comparison to males. Similarly, using LIWC2001 version, Newman, Groom, Handelman, and Pennebaker (2008) studied a corpus of 14,000 text files from 70 different studies to see gender differences in language use. They also found that women used more third person

singular pronoun. In the same vein, a corpus-based study of outsourced call centres discourse by Friginal (2009) confirmed that female callers used personal third person pronoun (*he/she*) more than male callers and claimed that the results were consistent with the previous research (Beeching, 2002; Argamon, Koppel, Fine, & Shimoni, 2003; and, Mills, 2003 as cited by Friginal, 2009, p.129). More recently, a study by Yu (2014) replicated the model of Newman, Groom, Handelman, and Pennebaker (2008) and analysed gendered language differences in political setting from a corpus of Congressional speeches delivered between 1989 to 2008. The study found that, in case of overuse of third person singular pronoun by women, the results of both the studies were consistent.

In contrast to these studies, some research has highlighted other interesting aspects of gender differences in the use of personal third person singular pronoun. For example, for a study of weblogs, Herring and Paolillo (2006) based their analysis on language features of men and women as hypothesised by Argamon, Koppel, Fine, and Shimoni (2003). They concluded that third person masculine was favoured by women while third person feminine did not show any significant gender differences. Similarly, a study by Ahmad and Mehmood (2015) applied CDA techniques to analyse parts of speech for examining differences in language use by men and women in a corpus of Pakistani English newspaper columns. The study found that men used more masculine third person singular (*he*) whereas women used more feminine third person singular (*she*).

In addition to these contrasting differences, some research has not found any significant differences in the overall use of third person singular pronoun by men and women. For example, a LIWC-based study by Lenard (2017) analysed a corpus of Congressional speeches for pronoun

use by men and women legislators and reported no significant differences in the use of personal third person singular (*he/she*) on the basis of gender.

The overall result of the present study for personal third person singular pronoun (*he/she*) and its variants were consistent with Lenard (2017). The results of Mann-Whitney U-test conducted for the present study revealed that there are no significant differences in men and women bloggers in the use of this pronoun (U = 969408; U (standardized) = -2.040; Expected value = 1014444.000; Variance (U) = 487386102.494; p-value (Two-tailed) = 0.0414).

Although, the hypothesis test for the present study showed that there is no significant difference in the use of personal third person singular pronoun by men and women bloggers, data for this pronoun was further broken up to see any subtle differences that did not surface in the results for the overall quantitative testing of the null hypothesis.

As Figure 5.6 indicates, it was observed that both men and women bloggers used, with greater frequency, their *own* gender-directed personal third person singular pronoun. As a result, men tended to use more third person singular pronoun *he, his, him, himself* and women scored higher on *she, her, hers, and herself.* This observation of the present study supports the findings of Ahmad and Mehmood (2015). This tendency in both the genders may be attributed to the socio-cultural make-up of Pakistan where a frequent reference to the opposite gender may not be socially desirable.

Another interesting observation was the consistency of men bloggers to overuse the contracted variants of auxiliaries with this pronoun even if the pronoun case referred to female gender (e.g. *she's*). This observation was consistent with previous observations of this study for contracted variants of auxiliaries with other pronouns.

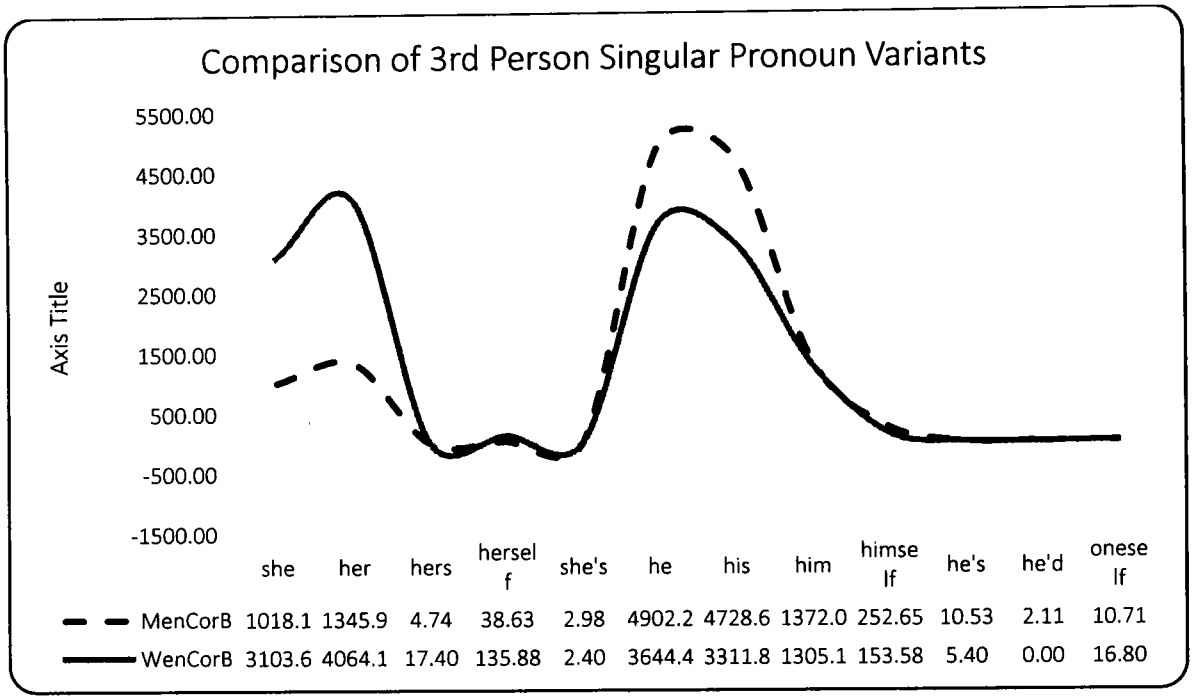


Figure 5.6. Personal third person singular pronoun variants (frequencies normalised to one million)

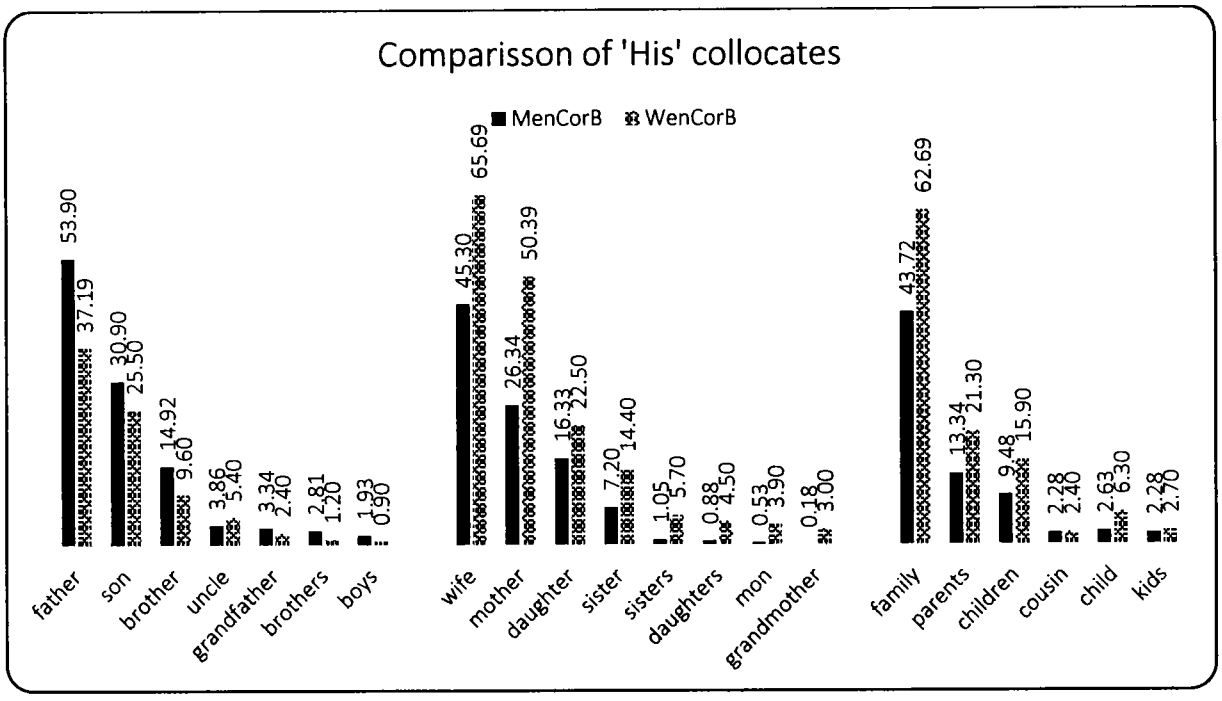


Figure 5.7. R1 noun-referents (masculine, feminine, neuter) to family members with 'His' in GenCorB (frequencies normalised to one million)

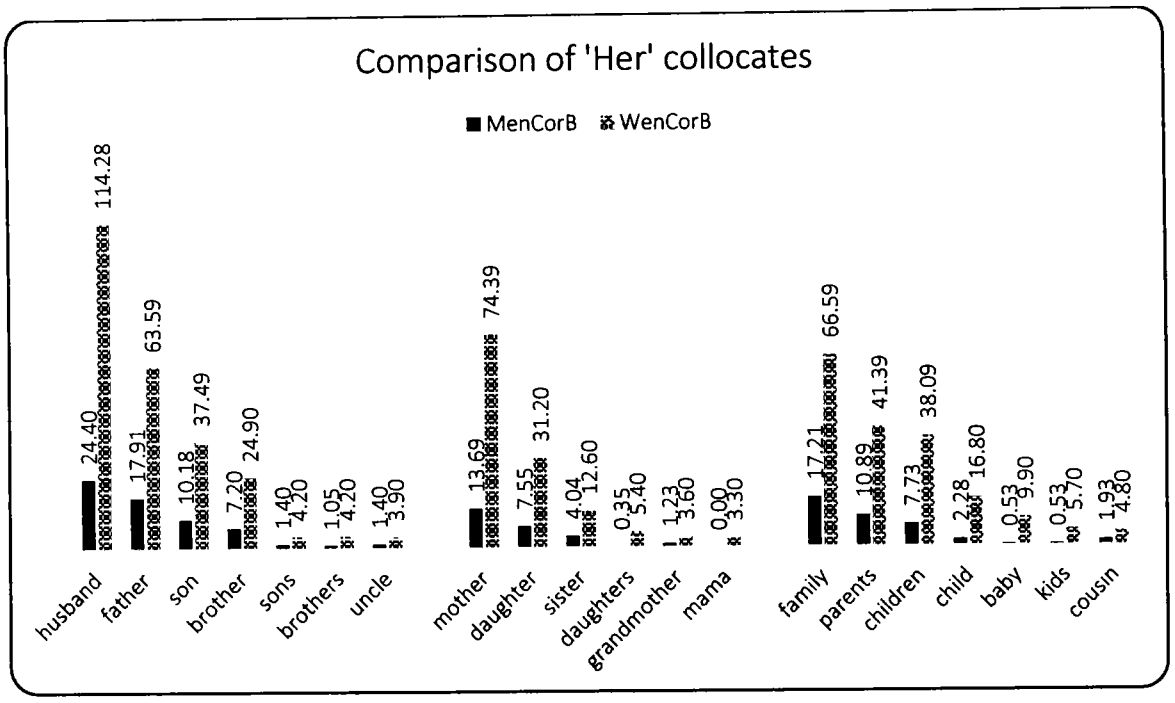


Figure 5.8. R1 noun-referents (masculine, feminine, neuter) to family members with 'Her' in GenCorB (frequencies normalised to one million)

After observing these patterns, the analysis was further zoomed in to see differences in the use of possessive cases of the personal third person singular pronouns (*his/her*). Since the analysis of all the immediate right collocates (R1) of these possessive cases was not possible, the examination was limited to the R1 collocates of the nouns that referred to family members. The analysis was based on grammatical gender sub-categories of nouns, i.e. masculine, feminine and neuter nouns, which yielded some interesting patterns as indicated by Figures 5.7 and 5.8.

A combined examination of the two Figures showed that, with the only exception of the noun 'uncle', men used more masculine nouns when talking about relations of other men (e.g. *his* father) but they used less feminine and neuter nouns when talking about relations of other men that they referred to with *his*. In contrast, women used more feminine and neuter but less masculine nouns when talking about relations of other men. They, however, used more masculine, feminine and neutral nouns when talking about relations of other women. This tendency, again, signal,

gender polarity in men and women, which can be attributed to the overall gender polarity that prevails in Pakistani society.

Another subtle difference was the collocation of the term 'grandfather' with the third person possessive pronoun *his/her*. Surprisingly, the noun 'grandfather' collocated with masculine third person singular possessive case *his* more in both MenCorB and WenCorB than with the feminine third person singular possessive case *her*. Thus, in both the corpora *his grandfather* occurred 27 times (19 times in MenCorB and 8 times in WenCorB) whereas *her grandfather* occurred 5 times (1 time in MenCorB and 04 times in WenCorB), which confirmed the findings of the present study that, in a patriarchal society like Pakistan, ancestral referents are generally used more for men than women and more by men than women.

Besides, some other interesting qualitative findings in the use of the terms *his grandfather* and *her grandfather* were also noted in both the corpora.

'grandfather' as an ancestor

ey have lived there for hundreds of years. He can call it	***his grandfather***	's house. It would be cruel to kick him
relative of **** It was ***** who told me that	***his grandfather***	's first name was ****, while
increasingly more radicalised with time. **** and	***his grandfather***	meanwhile had always stayed away from s
his father sent him, his mother and sister to Delhi, where	***his grandfather***	was very well-entrenched. His father stay
to see the temple and then left. Later, it was known that	***his grandfather***	**** was the head master of the
wrong time. Even though his statement made	***his grandfather***	instantly popular, it certainly cannot help his

grandfather as an authority

Shia. Years after, the grandson of this **** confessed that	***his grandfather***	wanted to bring down **** ra
ha prayer but stopped every passer-by to offer them water.	***His grandfather***	had mastered the craft of foretelling the loca
in their mind. For example, my cousin told me a saying of	***his grandfather***	after his Janaza prayer: "You cannot compare
legs and left him unharmed. He loved every part of that.	***His grandfather***	would say that the morning prayers would

'grandfather' as someone to be emulated

y, he likely feels compelled to deliver robust speeches like	***his grandfather***	. Alas, his voice box seems optimised for a s
and so has ****'s grandson, ****, who, like	***his grandfather***	, has begun to juxtapose leftist populist pos
eaking family in Karachi and grew up with stories of how	***his grandfather***	had fought against the British and was hanged
y being different. Although many debate that he is aping	***his grandfather***	**** in the slapstick comedy scenes, w
certainly did. He does have football blood after all. From	***his grandfather***	to his father to now the youngest ****, t
ly exudes charisma like his mother **** and	***his grandfather***	****, but also presides over a r

been better if he had opted to copy	***his grandfather***	, ****, rather than his mother.
out of him. Had **** copied	***his grandfather***	's style, he might have been able to
It seemed as if, in attempt to replicate	***his grandfather***	****, he forgot the obvious yet delicate
of Pakistan. Baby **** might dress like	***his grandfather***	in his signature awami shalwar suit with

'grandfather' as ordinary family elderly relation

Sohan Halwah, ****, however, asserts that	***his grandfather***	learned to prepare Sohan Halwah from Del
e to find out about visa procedures and eventually leads	***his grandfather***	to his Indian friend to end in an emotional
ing his speech everything was common between him &	***his grandfather***	**** not only this but
gs stand currently, he will not be present in Pakistan for	***his grandfather***	's death anniversary and this is something tha

Concordance 5.15 '*his grandfather*' occurrences in GenCorB

As shown in Concordance 5.15, it was observed that the use of *his grandfather* in both the corpora was mainly focussed on the role of this relationship as a source of lineage for the men spoken about. In this sense, *grandfather* was used as an *ancestor* to be recalled, an *authority* to be quoted, a *model* to be emulated and an *ordinary family elder*. In contrast, as indicated by Concordance 5.16, in all occurrences of *her grandfather*, the relationship of grandfather was presented as an *ordinary family elder* in both the corpora.

'grandfather' as ordinary family elderly relation

, where she spent most of her time, dancing to	***her grandfather***	playing his tabla. The 70-year-old music a
tie the scarf and she smiled at me and requested	***her grandfather***	to bring some more colorful headscarves fo
easily be characterized as the suburbs of Lahore.	***Her grandfather***	had died when she was very young, but she
that her mother's finesse had to be attributed to	***her grandfather***	alone. She knew but two things about that o
that she has no one to call 'papa' anymore since	***her grandfather***	had also passed away due to an illness a mont

Concordance 5.16 '*her grandfather*' occurrences in GenCorB

Correspondingly, the occurrence of *his/her* with grandmother were also studied. It was found that the word 'grandmother' occurred with *her* (7 times in MenCorB and 12 times in WenCorB) more in both the corpora in comparison to its occurrence after *his* (01 time in MenCorB and 10 times in WenCorB). Thus, taken together, *her grandmother* was used 19 times whereas *his grandmother* was used about half of its frequency, i.e. 11 times in GenCorB. Beside this quantitative analysis, the concordances of *his grandmother* and *her grandmother* were also studied and some subtle patterns were observed as indicated by Concordances 5.17 and 5.18.

'grandmother' as ordinary family elderly relation

he was born and to have a cup of tea with	***his grandmother***	. So, we went there to have a harmless cup of
bonds can blossom, such as the ones between Mustafa and	***his grandmother***	, and between **** and his mother. It all hinges
, namely the spouse in this case. Her not letting him attend	***his grandmother***	's funeral, must it be shelved as the alternate pers
boy, Faizan, was brought to the hospital where I work, by	***his grandmother***	. She complained that he suffered from severe diarr
e fire. His elder brother was injured. The deceased had told	***his grandmother***	a day before the accident: "Mein school toh khud
of. Said to be a particular childhood favourite of ****,	***his grandmother***	, the ****, made sure the chocolate biscuit cake
olunteers to accompany her to India, entirely out of love for	***his grandmother***	, albeit with some uncertainty. Being born and bred
re he visits the fields on the surrounding mountainside with	***his grandmother***	. Mohamed finally finds God in the last scene as h
oblivious of their father's plans to take him away and	***his grandmother***	is not strong enough to retaliate against her son.
arna dude: Who? Pappu Saeen? Pathan: Yes, Pappu Saeen,	***his grandmother***	, and the ****. Dharna dude: Dude, why
l Chávez Frías was born on July 28, 1954. He was raised by	***his grandmother***	after he and an older brother were placed with her

Concordance 5.17 '*his grandmother*' occurrences in GenCorB

'grandmother' as an authority

to go to the United States. ****, on the insistence of	***her grandmother***	, agrees to accompany ****. **** is still emo
until she is a young adult. Out of everything that	***her grandmother***	ever told her, one thing definitely stands out in
mbers the life lessons that her Uncle **** and	***her grandmother***	had taught her. In fact, both had told her to
s recipes for whitening her skin, provided to her by	***her grandmother***	, like turmeric or saffron, only then will she be te

'grandmother' as someone to be taken after

ine of **** and **** and inherited	***her grandmother***	, **** striking good looks. 7. ****
er mom. It would have been better if she resembled	***her grandmother***	." With these comments, you can't help thinking,

'grandmother' as a family elderly relation

, she was given permission to bring the remains of	***her grandmother***	****. She finally interred these remains at
shawl from the wooden box, a wedding gift from	***her grandmother***	. Like half a million displaced tribesmen, their to
Video blog: Stepping into the slums With	***her grandmother***	by her side, an innocent girl, ****, stared at me.
s born in, just outside of Johor town in Lahore. As	***her grandmother***	described the ordeal that she had to go through jus
ng fortunate and guilty at the same time. **** and	***her grandmother***	, regrettably, are not the only people struggling to
f this. One time, I assumed my friend's mother was	***her grandmother***	, by mistake of course, and you can imagine how of
ound. **** also shares a unique relationship with	***her grandmother***	, who is a rather spirited, graceful woman. She tells
as French and not Iranian. As she walks, she hears	***her grandmother***	ask in her usual soft voice that was she now
ep further when the wolf serves Little Red a plate of	***her grandmother***	's flesh, which the little girl eats. Cannibalism and
daughter returns as a beautiful young woman, filling	***her grandmother***	's heart with joy. In truth, the whereabouts of the
be Anastasia, but the real Anastasia did not return to	***her grandmother***	and there was no 'happily ever after'. The cartoon

'grandmother' as a symbol of tradition

She soon revolted against tradition and questioned	***her grandmother***	on why she kept the utensils of her father's
re, and their tailor stitched it. Meeha was decked in	***her grandmother***	's jewellery on her big day. Her sister supervis

Concordance 5.18 '*her grandmother*' occurrences in GenCorB

It can be noted from the above concordances that for the words 'grandfather' and 'grandmother', the concordance patterns of third person singular pronouns *his/her* presented a contrasting pattern when the third person singular possessive cases occured with nouns of opposite gender, the noun was used in one sense only. Thus, the word 'grandmother' occurred with *his* in one ordinary sense, i.e. a family elderly relation, as did the word 'grandfather' with *her*. However, where the third person possessive case was used with nouns of the same gender, the noun was used in divergent connotations. Thus, the words 'grandfather' and grandmother' occurred with *his* and *her* respectively in some additional senses besides their ordinary sense of 'an elderly family relation'.

A subtle difference was also noted in the use of the words 'grandfather' and 'grandmother' in another sense.. The word 'grandfather' was used in the sense of a family ancestor but the word 'grandmother' was not used in this sense. Moreover, the word *his grandfather* was used for men in the sense of a model to be copied or followed in some skills or style (e.g. oratory or dress up) or socio-political orientation. In contrast, the word *her grandmother* was used in the sense of someone to be taken after by women for qualities that related to physical appearance.

5.4.8 Third Person Plural Pronoun (*they*)

Research in the past on difference between men and women in the use of the third person plural pronoun (*they*) was fuzzy. These studies did not pay exclusive attention to the use of third person plural pronoun (*they*) and its variants. Most of these studies analysed this pronoun alongside other pronouns. For instance, in an effort to investigate simple lexical and syntactic features to infer authors' gender of unseen formal written texts, Koppel, Argamon, and Shimoni (2002) employed machine learning algorithms on a corpus of 566 documents from the BNC to construct

models for performing such tasks. They observed that except for the third person singular pronoun *he*, which was used with an equal frequency by men and women writers, women used all other pronouns with greater frequency. Later, Argamon, Koppel, Pennebaker, and Schler (2007) analysed over 140 million words of English text drawn from randomly selected blogs of men and women. They applied factor analysis and machine-learning techniques to see language differences on the basis of gender and age. They also made a general conclusion about the pronoun overuse by women. Similarly, Newman, Groom, Handelman, and Pennebaker (2008) examined gender differences in language using LIWC2001 version on standardized style-based and content-based categories to analyse a corpus of more than 14,000 text files from 70 separate studies. They also treated third person singular and plural forms together and concluded that women used more third person pronouns.

Only recently some research exclusively treated third person plural pronoun. For example, Argamon, Koppel, Fine, and Shimoni (2003) explored possible variation between male and female writing styles in Modern English, by studying a large subset of the British National Corpus (BNC) covering a range of different genres. The documents included in this study were all articles and books intended for an unseen audience. The study concluded that, in comparison to non-fictional texts, male authors used more plural pronouns (*we*, *us*, *they*, *them*) in fictional texts. Another study was conducted by Bell, McCarthy, and McNamara (2006), which investigated gender differences in language use in the light of the *biological* versus *social constructionist* theories of the origin of gender. Using LIWC for text analysis, the research examined language differences between men and women within the context of marital conflicts. The study predicted the use of more social words by women, among which was included the pronoun *they*. However, this corpus analysis yielded no significant differences in the use of the social words (pronoun *they*) by men and women.

A latest study by Lenard (2017) aimed to see gender differences in the personal pronouns use. The study examined a corpus of uninterrupted speeches (672 by women and 3,655 by men) delivered in the 113th United States Congress. The findings of this study also revealed no significant difference between male and female US congress members in the use of the third person plural pronoun and the results were consistent with those of Bell, McCarthy, and McNamara (2006).

The results of the present research are also consistent with those of Bell, McCarthy, and McNamara (2006), and Lenard (2017). The Mann-Whitney U-test conducted for present study also did not record any significant difference between men and women bloggers in use of third person plural pronoun (*they*) and its variants (U = 939217.500; U (standardized) = -3.405; expected value = 1014444.000; Variance (U) = 488097224.313; p-value (two-tailed) = 0.0007).

Once the hypothesis test for the present study showed that there was no significant difference in the use of third person plural pronoun by men and women bloggers, data for this pronoun was further dissected to see any subtle differences that did not surface in the results for the overall quantitative testing of the null hypothesis.

As Figure 5.9 indicates, an interesting observation was the consistency of men bloggers to overuse the contracted forms of the auxiliaries with this pronoun in comparison to women bloggers. This observation was internally consistent with previous observations of this study for other personal pronouns.

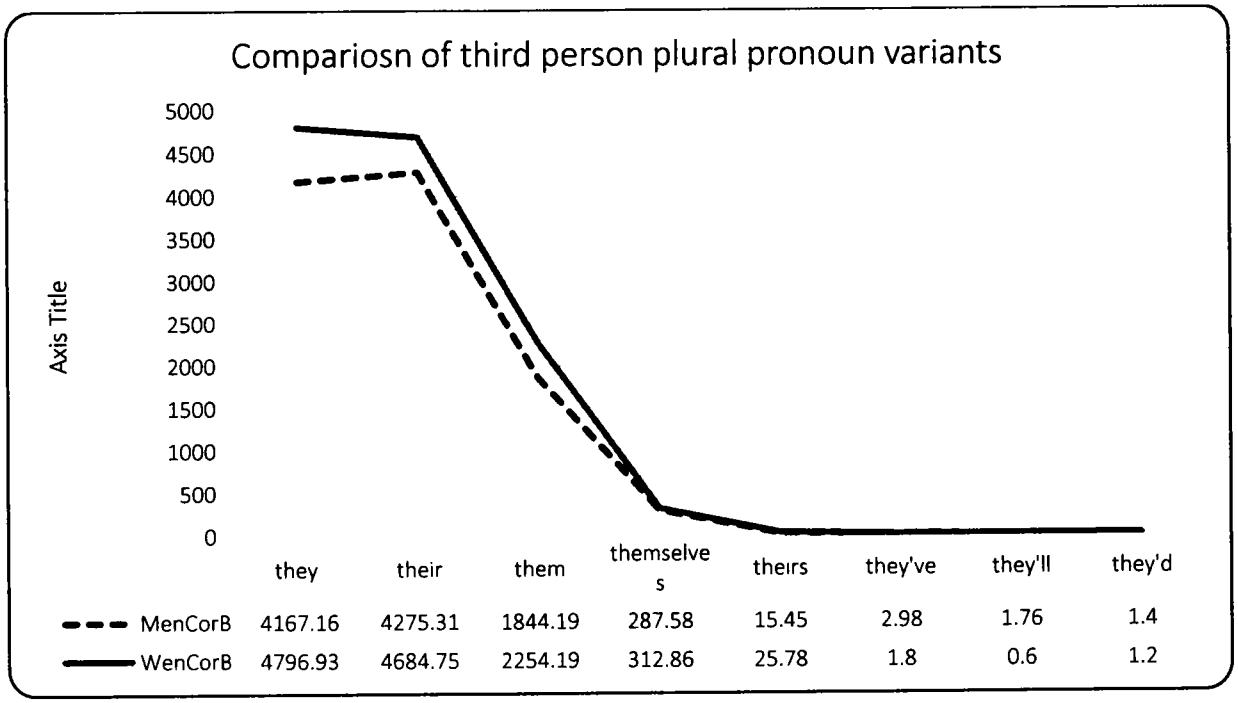


Figure 5.9. Comparison of third person plural pronoun variants

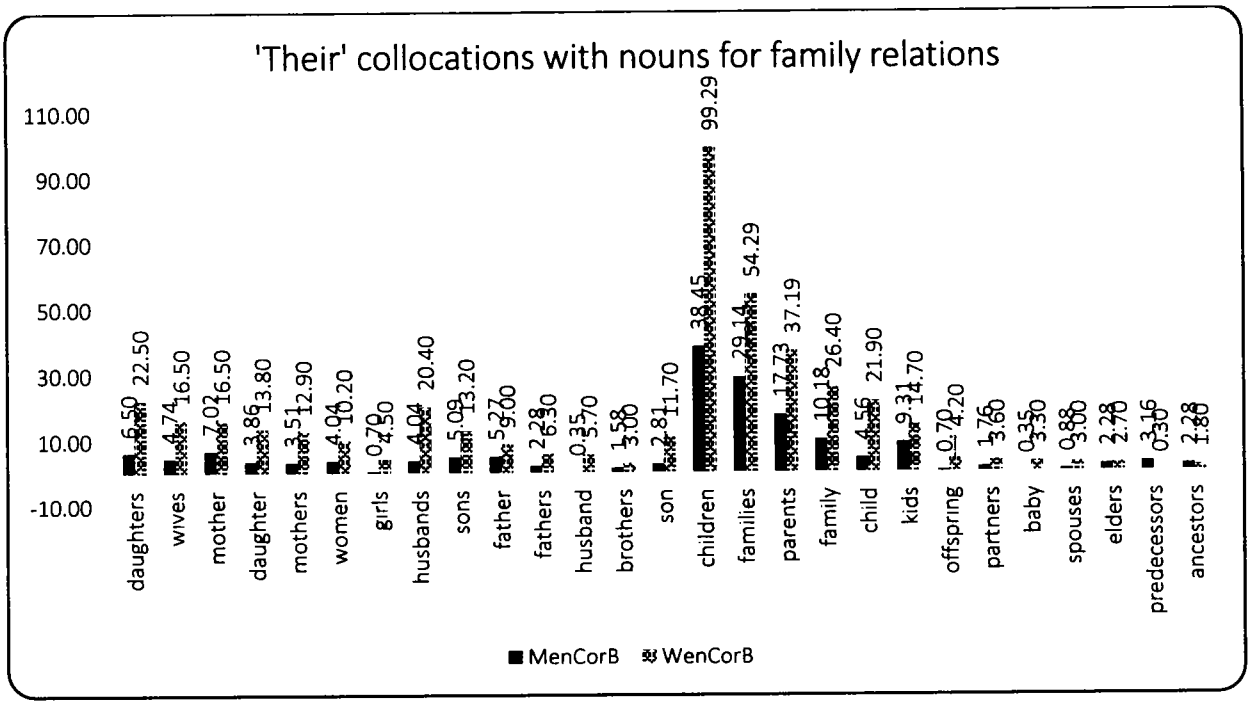


Figure 5.10. 'Their' collocations with nouns of family members (frequencies normalised to one million)

After observing this pattern, the analysis was further zoomed in to see differences in the use of possessive cases of the personal third person plural pronouns (*their*). As previously conducted for possessive cases of other personal pronouns, the examination was limited to the R1 noun collocates of *their* that referred to family members. The analysis was based on the grammatical gender sub-categories of nouns (i.e. masculine, feminine and neuter nouns), which yielded some interesting patterns as indicated by Figure 5.10.

It was observed that women bloggers collocated nouns for all family relations more than men bloggers except the last two instances of *their predecessors* and *their ancestors*. This tendency in men bloggers was consistent with their overuse of the nouns *elders*, *grandfather*, and *ancestors* with the first person singular possessive case *my* and *grandfather* with *his*. This consistent trend corroborates the findings of the present study that men tend to focus more on nouns that refer to ancestry and heads in family lineage beyond parental level.

The data was also scrutinised for the use of the third person subjective case (*they)* by men and women to see who they refer to with this pronoun. As Malone (1997) observes, third person pronouns are used for creation of an object that a speaker or writer desires to talk about in a particular manner. Unlike the first and second person pronouns, third person pronouns point to who we are not. As Sacks (1992) points out, third person plural (*they*) and its variants can be used for comparison between self and others, as institutional markers, or as agents of some organisation. In the present study, the data of GenCorB was analysed for the aspect of 'comparison between self and others' in two contexts: the construction of the 'other' in an *affiliative* and *oppositional* relationship. For this purpose, the context of *they* was studied through a systematic random sample of 100 concordances each from MenCorB and WenCorB.

In an *affiliative* context, men and women make a positive evaluation of '*they*' and the effect is to create an affiliation between the group referred to by '*they*'. Conversely, in an *oppositional* context men and women make a negative evaluation of '*they*' and the effect is to create a disaffiliation between them and the group referred to by '*they*'. In the case of a neutral context, men and women evaluates '*they*' neither positively nor negatively; that is, the bloggers of both the genders take up a neutral position in relation to the group being referred to as '*they*'. The effect of this is to create a neutral relationship with '*they*'. These three contexts depend on factors, such as the relationship that the bloggers create with the 'other' at a particular time and the issue being discussed.

In *affiliative* context, '*they*' is used for groups with whom men and women like to be affiliated or on whose behalf men and women want to talk but who are still included in the category of '*they*'. The instances represented by Concordances 5.19 and 5.20 show how men and women take up a positive position towards others invoked by 'they.'

Someone against unjust system

weren't protesting against the murderer, who is an individual; ***they*** were expressing their contempt and hatred towards a system

Concordance 5.19 Affiliative context of 'they': MenCorB

animals

ed to have someone speak on their behalf. That is one thing ***they*** are dependent on us humans for and we should not let them down.

children

news reports that poured in the figures of causalities as though ***they*** were mere numbers and not children. Without confirmed reports, inf
ght them to talk about or share these things with a trusted adult. ***They*** will suffer in silence for years with the effects of it continuing in their

same sex group

dress loosely, they are "too masculine" and must be a "dyke". If ***they*** wear makeup or fancy clothes, they are trying to "impress other people
musing about the "ripe" age for a girl's marriage, claiming that ***they*** become "too fast and out of control" or "get Western ideas" if their ma
public: that no matter what these women have achieved in life, ***they*** are undisciplined – because they smoke. Even when it comes to divorc
are of their rights and are more liberated than before? Why are ***they*** not raising their voice against domestic abuse? And also, are men really
Express your love and appreciation to them and remember that ***they*** are your better half, your companions in Jannah and half of your own
you make it your onus to find them worthy spouses because ***they*** deserve it, for the hard work and dedication they have mustered to
not give up in the face of any opposition. Not too long ago, ***they*** stood along men to demand independence from the British rule; they
found ways to avoid the mandatory chaddar; how and when did ***they*** give in to the dictates of a society that judges them by what they wear

skewed representation for men compared to women in Pakistan. ***They*** are such evil "ladies" because they refuse to silently obey and follow

victimised group

displaced by the horrific incident that took place in Mehrabadi. ***They*** are victims of the madness of intolerance that overtakes Pakistan in the

Concordance 5.20 Affiliative context of 'they': WenCorB

It can be noted from the above concordances that both men and women used 'they' in affiliative context for groups who suffer from some kind of injustice. In particular, unlike men, women bloggers appear to be concerned more with the social status and problems faced by the members of their own sex group. Also, they seem to sympathise with members of other weak and dependent groups like animals and children.

The pronoun 'they' has also been used both by men and women bloggers to distinguish themselves from others in oppositional contexts, i.e. by taking a negative position against others. In this sense, '*they*' is used to create an '*us*' and '*them*' distinction as exemplified by Concordances 5.21 and 5.22.

the law breaker

s locality is home to illegal Afghan refugees and ***they*** must go back and (ii) the locality is breeding grou

the cruel

spect for these creatures. During political rallies, ***they*** will often carry live lions and tigers across town in

the corrupt

p fighting for control of economic resources, so ***they*** can mint money off the public. But the cruelest of

e who do not provide an affidavit confirming that ***they*** are not dual nationals should be disqualified. T

the personally opposed

haters went ballistic this year more than ever. And ***they*** left their traces on my Facebook page as well sinc

Concordance 5.21 Oppositional context of 'they': MenCorB

the privacy intruder

ink that child birth is their key to a lock with which ***they*** have now access to the doors of your privacy. They feel

le ones who inform you an hour before coming that ***they*** 're dropping in for a cup of 'chai'! However, many are

the violator of women's rights

g some crazy extremist had said somewhere where ***they*** were still flogging women in the streets. Unfortunately,

her. December 20, 1995 They're beating her again. ***They*** 're punishing her for not succumbing to their "pure" de

te" and "face" everything no matter what happens. ***They*** embed the thought in these girls' mind that tolerating ab

wheels will always look down upon women drivers. ***They*** wear a disgruntled expression on their face with their ey

the opposite sex

Men in their 20s continue to live in a bubble where ***they*** think they will be young forever. But in their early 30s, t
th men stayed in a promiscuous rut for a while until ***they*** fell for, incidentally, the same woman. Although Brand

Concordance 5.22 Oppositional context of 'they': WenCorB

As shown by the above concordances the use of '*they*' in oppositional context creates a dichotomy. Men tend to use oppositional context for those who violate law, are cruel and corrupt, or engage in personal opposition. In the personal *oppositional* context, men create a dichotomy not between '*us*' and '*them*' but between '*me*' and '*them*'. The focus of men's *oppositional* context seems to pivot around socio-political and personal issues. In this context, men seem to speak either on their own behalf or on behalf of the society as a whole.

In contrast, the *oppositional* dichotomy created by women is between '*us*' and '*them*'. Unlike men, women do not use, *oppositional* context in a personal oppositional sense. Moreover, in the instances given in Concordance 5.22, women create a dichotomy between them and others on the basis of privacy intrusion, violation of their rights and their relationship with the opposite sex group. Thus, the focus of their oppositional context seems to pivot around the problems faced by their own sex group. In this context, women do not seem to speak on behalf of the entire society but on behalf of other women, i.e. their own sex group.

To sum up the discussion on the use of the third person plural pronoun *they*, the results of the present research are consistent with those of Bell, McCarthy, and McNamara (2006), and Lenard (2017). The present study did not find any statistically significant difference between men and women in the use of this pronoun. Some of the findings of the present study for other personal pronouns were also corroborated by the use of *they*. For example, men were found to overuse this pronoun with the contracted form of auxiliaries. Also, they were found to overuse words like *ancestors* and *predecessors* with the possessive case of this pronoun too. In the same vein, women

were found to overuse words that refer to kinship after the possessive case of this pronoun. Both the genders were also found to use the pronoun *they* to identify with group(s) they belonged to or liked to belong to (i.e. the use of they in *affiliative* context) as well as for group(s), which they did not belong to or did not like to belong to (i.e. the use of 'they' in *oppositional* context).

5.4.9 Impersonal Pronouns

This category of LIWC2015 includes third person impersonal pronoun *it* as well as indefinite and relative pronouns. This category of pronouns, unlike personal pronouns, has been assigned limited attention by the previous research on gender differences in language use. The few studies that have analysed gender differences in impersonal pronouns include a study by Raumolin-Brunberg (1998) who examined significant gender differences in the pronominal changes at different periods in the sample language text caused by a Civil War fought in the 17th century period. The study found that women used the possessive adjective *its* more in the third period observed and *who* after the war in comparison to men. In contrast, *men* were found to have overused compound impersonal pronoun variants with *-body* suffix in the third period but the same variants were used more by women in the first period. Men were also found to overuse impersonal pronoun variants ending in *-one* suffix in the first period but in the third period both the genders used them with almost the same frequency. The study concluded that women accepted new alternatives and led the language change. Using LIWC as a tool, another study by Brownlow, Rosamond, and Parker (2003) analysed gender differences in language use in televised interviews. The study concluded that men used more impersonal pronoun in comparison to women. Similarly, a LIWC-based study by Nagarajan and Hearst (2009) examined gender-based language differences in online dating sites. The study was particularly focused on the section where the members

described themselves. The study found that men used more impersonal pronouns in comparison to women. The results of a more recent study by Lenard (2017) on gender differences in the language of political speeches also showed that men used more impersonal pronouns.

The results of the present study, however, were not consistent with those of Brownlow, Rosamond, and Parker (2003), Nagarajan and Hearst (2009), and Lenard (2017). The Mann-Whitney U-test conducted for the present study found that women bloggers used more the impersonal pronoun in comparison to men bloggers (U = 892222.500; U (standardized) = -5.532; Expected value = 1014444.000; Variance (U) = 488111498.348; p-value (Two-tailed) = < 0.0001.

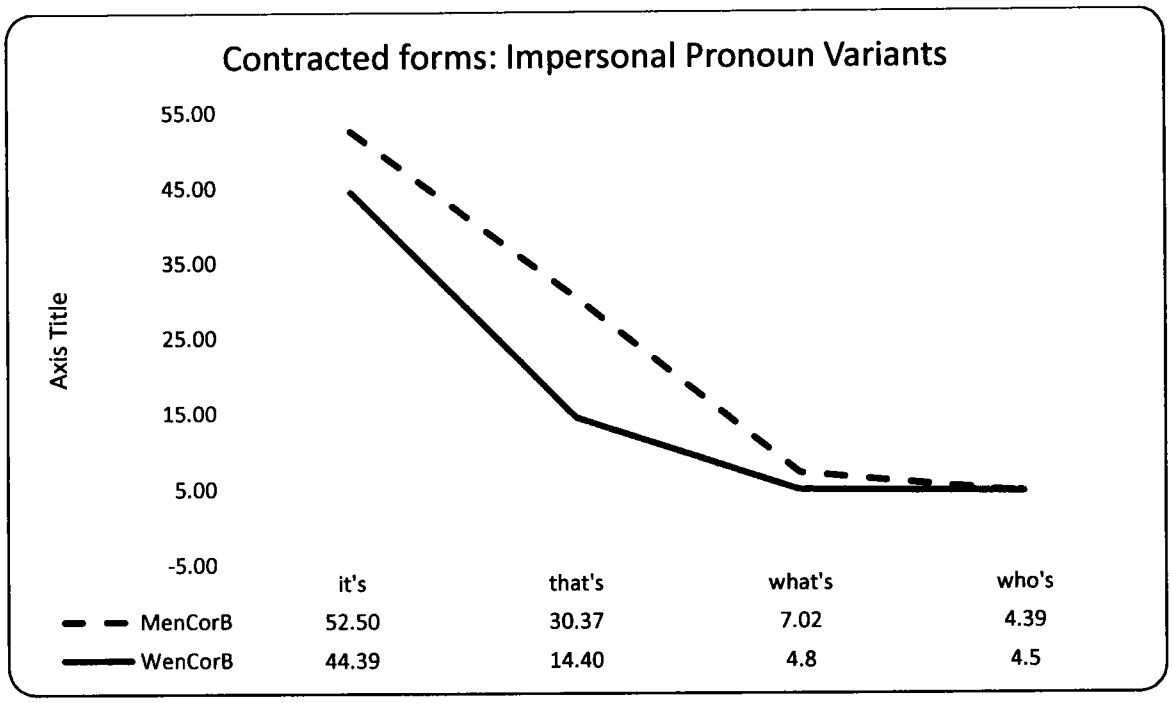


Figure 5.11. Contracted forms: Impersonal Pronoun Variants (frequencies normalised to one million)

Besides this hypothesis test, the data was further probed to find any differences in the use of impersonal third person pronoun variants with the contracted forms of auxiliaries. As shown in Figure 5.11, the results were consistent with those found for other pronoun variants in the present study: men tended to overuse impersonal pronoun with contracted forms of auxiliaries.

5.4.10 Articles

It may seem uninteresting to study articles as their use in language is considered essential and is guided by English syntax. Although, articles may not represent some particular theme in a text, they are important to study because they keep company with nouns whose use, in turn, shows a speaker's or writer's ability to categorise things (Lenard, 2016).

A number of studies in the past have examined the use of articles by men and women, both in spoken and written discourses. For instance, about six decades ago, in their study on verbal recording of five-minute speech samples, Gleser, Gottschalk, and Watkins (1959) reported that men used articles more than women. Their study also showed a positive correlation between the use of more articles to the speaker's higher intelligent quotient (IQ). Later, other studies on gender differences in language use yielded consistent results like Mulac and Lundell (1980) on description of photographs; Mehl and Pennebaker (2003) on college students' conversation; Brownlow, Rosamond, and Parker (2003) on television interviews; Argamon, Koppel, Fine, and Shimoni (2003) on formal written texts; Schler, Koppel, Argamon, and Pennebaker (2005) on blog writing; Newman, Groom, Handelman, and Pennebaker (2008) on 14,000 text samples; Kapidžić and Herring (2011) on self-representation in chatrooms; Flekova and Gurevych (2013) on author-profiling in social media; and, Yu (2014) on political speeches.

The findings of two studies in the literature examined, however, were not consistent with the previous research. One was the research of Ludu (2014) on Twitter data, which did not find any significant association between gender and the use of articles. The second study was that of Lenard (2016) on Congressional speeches, which did not find gender differences in article use. It is interesting to note that both of these studies did not find any significant difference in the use of articles by men and women; however, none of the studies found that women used more articles.

The findings of the present study confirmed the previous research. Mann-Whitney U-test conducted for the present study showed that men used more articles in comparison to women bloggers (U = 1296211.500; U (standardized) = 12.754; Expected value = 014444.000; Variance (U) = 488113859.099; p-value (Two-tailed) = < 0.0001).

Since the use of articles is related to nouns, a conclusion may be drawn that men talked more about specific entities to which nouns in the text refer. As Pennebaker (2013) pointed out, men talk in clear-cut ways about things as their thinking pattern is more categorical in comparison to women's; hence, men overuse articles, which occur with nouns.

5.4.11 Prepositions

Previous research on gender differences in language use has approached the use of preposition by men and women from two major perspectives, i.e. error analysis and difference in preposition frequency. From the error analysis perspective, Shofwan, Musli, and Sugeng (2013) found that, compared to male students, female students made fewer errors in preposition use. Hence, they claimed that women were more confident users of prepositions. Based on past research findings, Saeed, Hussain, and Fatima (2015) hypothesised that men committed fewer errors in prepositions. For this purpose, they studied 26 ESL students. They, however, found that while female committed more overall errors in writing, they also committed more prepositional errors in cases of 'place' and 'time' prepositions. Male, on the other hand, were found to commit errors in 'direction' prepositions.

From the latter perspective, previous studies yielded contradictory results. For instance, Koppel, Argamon, and Shimoni (2002) observed that women used the preposition *for* and *with* more frequently than men. In contrast, they reported, that men used the preposition *of* more

frequently than women but men used the set of all other prepositions with the same frequency as women do. Similarly, Argamon, Koppel, Pennebaker, and Schler (2007), who studied blog posts, reported prepositions to be a strong indicator of men's style. This tendency in men's style was also confirmed later by Newman, Groom, Handelman, and Pennebaker (2008) who studied a large sample of 14000 texts. While these studies reported a significant difference in the use of prepositions by men and women, Bamman, Eisenstein, and Schnoebelen (2014) as well as Baumann, Krasnova, Veltri, and Yusni (2015) did not find any significant difference in Twitter texts. Thus, previous research either found no differences between genders in the overall use of prepositions or found that men used more prepositions.

The findings of the present study are consistent with the previous research that found that men used more prepositions in comparison to women. The Mann-Whitney U-test conducted for the present study showed that men used more prepositions in comparison to women bloggers (U = 1127191.500; U (standardized) = 5.103; Expected value = 1014444.000; Variance (U) = 488113022.675; p-value (Two-tailed) = < 0.0001).

Table 5.1

Prepositions used more by men (frequencies normalised to one million)

S No	Prep	Fr. MenCorB	Fr. WenCorB	S No	Prep	Fr. MenCorB	Fr. WenCorB
1	of	30202.10	27382.89	19	among	267.74	194.07
2	in	20719.56	18472.63	20	including	234.03	188.37
3	on	6789.65	6723.26	21	above	186.28	158.08
4	as	6753.30	6440.10	22	near	122.90	122.38
5	from	4258.45	3994.54	23	throughout	117.98	97.79
6	by	4896.30	3934.55	24	ahead	87.26	77.39
7	into	1366.64	1347.11	25	unlike	92.53	60.29
8	after	1629.64	1333.61	26	below	68.65	59.69
9	over	1288.33	1237.63	27	besides	64.79	45.89
10	against	1344.34	816.79	28	Via	51.97	44.99

Table 5.1

Prepositions used more by men (frequencies normalised to one million)

11	since	679.63	580.72	29	plus	47.05	36.29
12	between	652.07	544.42	30	versus	43.72	29.10
13	under	602.21	493.43	31	Vs	47.40	21.60
14	during	689.46	454.44	32	atop	8.25	7.50
15	towards	415.40	381.55	33	minus	7.90	6.30
16	within	308.48	281.36	34	excluding	7.02	2.70
17	despite	281.09	272.36	35	hereafter	4.92	2.10
18	along	272.84	255.26	36	unto	4.92	1.50
				37	thru	0.70	0.00

Table 5.2

Prepositions used more by Women (frequencies normalised to one million)

S No	Prep	Fr. MenCorB	Fr. WenCorB	S No	Prep	Fr. MenCorB	Fr. WenCorB
1	to	27285.53	28787.90	20	outside	166.97	212.67
2	for	8972.86	9367.70	21	inside	140.81	197.37
3	with	6648.49	7010.62	22	till	155.03	155.38
4	at	3952.61	3989.15	23	beyond	136.07	151.48
5	about	2036.44	2585.64	24	amongst	118.86	119.38
6	out	1996.76	2230.49	25	regarding	89.37	104.99
7	up	1866.13	2170.50	26	except	79.18	90.89
8	than	1382.61	2132.10	27	unless	79.71	86.69
9	through	798.32	1398.71	28	onto	48.11	67.19
10	around	637.50	814.09	29	amidst	28.97	45.89
11	before	723.70	763.99	30	beside	15.80	25.20
12	down	613.44	746.30	31	toward	18.79	19.50
13	off	594.13	650.01	32	beneath	12.47	19.20
14	without	544.44	599.92	33	amid	10.89	17.40
15	away	495.11	595.72	34	underneath	10.89	13.20
16	across	336.22	567.82	35	respecting	7.55	11.10
17	behind	303.56	346.15	36	sans	6.50	8.10
18	until	244.04	312.26	37	abt	0.00	2.10
19	upon	255.81	269.66	38			

Besides rejecting the null hypothesis on this variable and finding that men used more preposition overall in comparison to women, the present study also tried to find out, which particular prepositions were used more by men and women. As listed in Tables 5.1 and 5.2, there

were some prepositions used more by men and others by women. Some of these results were consistent with the previous research. For instance, the present analysis found that the preposition *of* was used more by men whereas the prepositions *for* and *with* by women, which was in agreement with the findings of Koppel, Argamon, and Shimoni (2002).

5.4.12 Auxiliary Verbs and Common Verbs

LIWC2015 categorises auxiliary verbs and common verbs separately. However, since previous research on language and gender differences, as given below, has discussed the two variables together, this study also groups them in this section for discussion.

Among the earliest studies on auxiliaries and common verbs accessed for discussion here, a research was conducted by Gleser, Gottschalk, and Watkins (1959), which found that, compared to men, women used more auxiliary verbs and common verbs when they described some dramatic events in their lives. Later, McMillan, Clifton, McGrath, and Gale (1977) tested two hypotheses of Key (1975), and Lakoff (1975). One of the hypotheses was that women use more linguistic categories showing uncertainty. For this purpose, the study noted the frequency of verbs and modal verbs in recorded interactions between same-sex and mixed-sex groups, which were used for expression of uncertainty. The study also concluded that women used more verbs and auxiliaries in comparison to men. In studies using CL techniques, the results of Biber, Conrad, and Reppen (1998), Mulac, Seibold, and Farris (2000), and Pennebaker, Mehl, and Niederhoffer (2003) also showed that women used more verbs and auxiliary verbs, especially the auxiliary verb *could*, in comparison to men. The higher rate of *could* in women's language was explained as an indicator of their tentative style. Comparable results were reported by studies such as the analysis of author profiling by Argamon, Koppel, Pennebaker, and Schler (2007), 14000 text sample by Newman, Groom, Handelman, and Pennebaker (2008), students' letters in Iranian EFL context by Hamdi

and Dabaghi (2012), and social media contents by Schwartz, et al. (2013). The studies conducted by Yu (2014), and Lenard (2016), however, yielded contradictory results. Both the studies analysed political speeches and found that male politicians used more verbs in comparison to their female counterparts.

The results of the present study were consistent with the research that reported women as using more verbs and auxiliary verbs. The Mann-Whitney U-test conducted for the present study showed that women used more auxiliary verbs in comparison to men bloggers (U = 930116; U (standardized) = -3.817; Expected value = 1014444.000; Variance (U) = 488113344.817; p-value (Two-tailed) = 0.000). The same test also found that women used more verbs in comparison to men (U = 773889; U (standardized) = -10.888; Expected value = 1014444.000; Variance (U) = 488115209.926; p-value (Two-tailed) = < 0.0001).

A possible explanation of women's overusing verbs and auxiliary verbs may be what Pennebaker (2013) has mentioned as the power link. He argued by quoting studies (e.g. Galinsky, Magee, Inesi, & Gruenfeld, 2006; Galinsky, Magee, Gruenfeld, & Whitson, 2008) that powerful people make decisions, are task-oriented and pay attention to tasks, not people. Most tasks, in turn, require clear understanding of objects, events or features that are essential for the task, which are reflected by use of articles and nouns. Those with less power, in contrast, pay attention to others and themselves by using more pronouns and verbs.

5.5 Other Grammar

5.5.1 Common Adverbs

The interest of linguists in the study of adverbs to explore gender differences probably began with Lakoff's (1975) hypothesis that women used more intensifying adverbs than men. More recently, this specific feature of language gained renewed attention from linguists but thier

studies yielded contradictory results. For instance, research by McMillan, Clifton, McGrath, and Gale (1977), Biber, Conrad, and Reppen (1998), Mulac, Seibold, and Farris (2000), Mehl and Pennebaker (2003), Newman, Groom, Handelman, and Pennebaker (2008), and Jieun and Jae-Woong (2009) concluded that women used more adverbs in comparison to men. On the other hand, studies like Xiufan (2013), Yu (2014), and Lenard (2016) found that men used more adverbs in comparison to women. Yet some other studies reported no significant gender differences in the use of adverbs, e.g. Zaini, Hazirah, Saadiyah, and Kemboja (2012), and Hanafiyeh and Afghari (2014).

The results of the present research are consistent with those which found that women used more adverbs than men. The Mann-Whitney U-test conducted for this study showed that there is a significant difference between men and women bloggers in the use of adverbs (U = 810851; U (standardized) = -9.215; Expected value = 1014444.000; Variance (U) = 488111059.728; p-value (Two-tailed) < 0.0001).

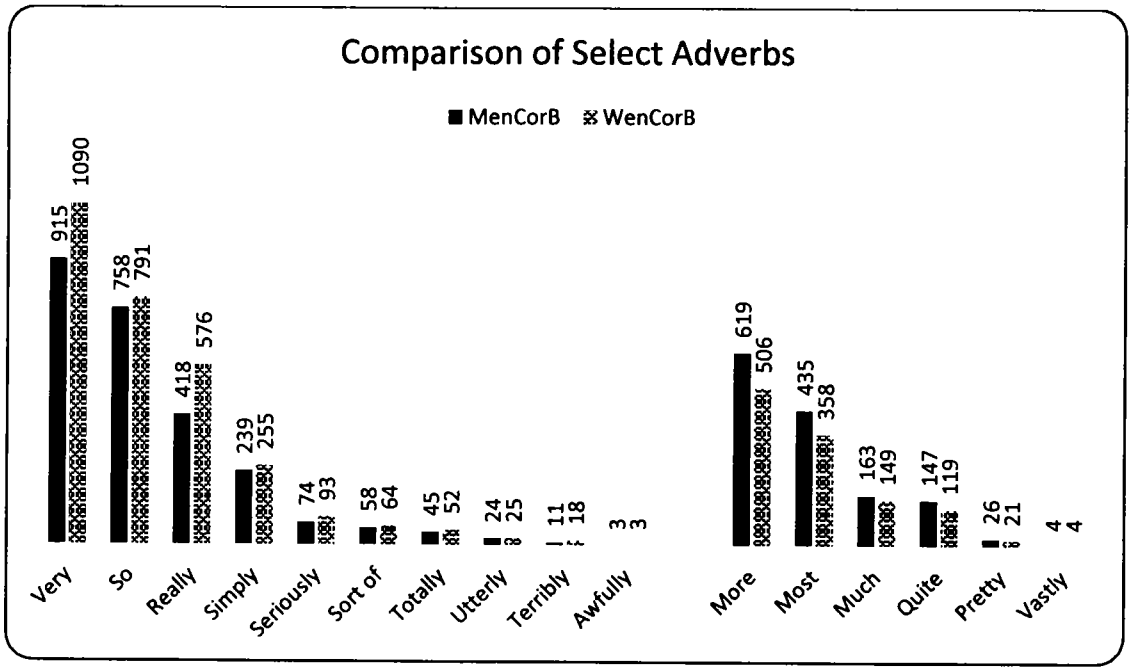


Figure 5.12. Adverbs (frequencies normalised to one million)

Table 5.3

Results of adverb use compared to previous research

S No	Adverb	Jeon & Choe (2009)	Xiufang (2013)	Lenard (2016)	Present Study
1	awfully	-	Women	-	Consistent
2	more	Men	-	Men	Consistent
3	most	Men	-	Men	Consistent
4	much	Men	-	Men	Consistent
5	pretty	-	Women	Men	Men
6	quite	Men	Women	Men	Men
7	really	-	Men	Men	Women
8	seriously	-	-	Men	Women
9	simply	-	-	Men	Women
10	so	Men	Women	Women	Women
11	sort of	Men	-	Men	Women
12	terribly	-	Women	Not significant	Women
13	totally	-	-	Men	Women
14	utterly	-	Men	Men	Women
15	vastly	-	Women	-	Consistent
16	very	Men	Men	Equal	Women

After finding the overuse of adverbs by women through hypothesis test, the use of select adverbs by the genders was confirmed. For this purpose, the adverbs mentioned by a few recent studies were analysed as shown in Figure 5.12.

It was found that women used *very*, *so*, *really*, *simply*, *seriously*, *sort of*, *totally*, *utterly*, *terribly*, and *awfully* more than men. In contrast, men were found to overuse *more*, *most*, *much*, *quite*, *pretty*, and *vastly*. Results for these adverbs were also compared to their respective source studies to see consistency in their use as shown in Table 5.3. The overuse of *awfully* by women and *more*, *most* and *much* by men was consistently found across studies. Results for the remaining adverbs were contradictory.

5.5.2 Conjunctions

Both in spoken and written discourse, conjunctions play a key role in making sentences more complex. Thus, they are good indicators of a speaker's or writer's language style (Lenard, 2016). Research on gender differences in language use has yielded contradictory results about use of conjunctions by men and women. For instance, studies by McMillan, Clifton, McGrath, and Gale (1977), Biber, Conrad, and Reppen (1998), Mulac, Seibold, and Farris (2000), and Mehl and Pennebaker (2003) found that women used more conjunctions than men, particularly the use of *but*, which was interpreted as an indicator of women's tentativeness in style. In contrast, Lenard (2016) found that men used more conjunctions than women in the corpus of political speeches that was analysed. Yet other research studies (e.g. Vali & Kianiparsa, 2010; Shofwan, Musli, & Sugeng, 2013) on error analysis of EFL students' writing on the use of conjunctions by the two genders reported no significant difference.

The results of the present study were consistent with the previous research, which found that women used more conjunctions in comparison to men. The Mann-Whitney U-test conducted for the study indicated a significant gender difference in the use of conjunctions (U = 808635.500; U (standardized) = -9.315; Expected value = 1014444.000; Variance (U) = 488110649.131; p-value (Two-tailed) = < 0.000).

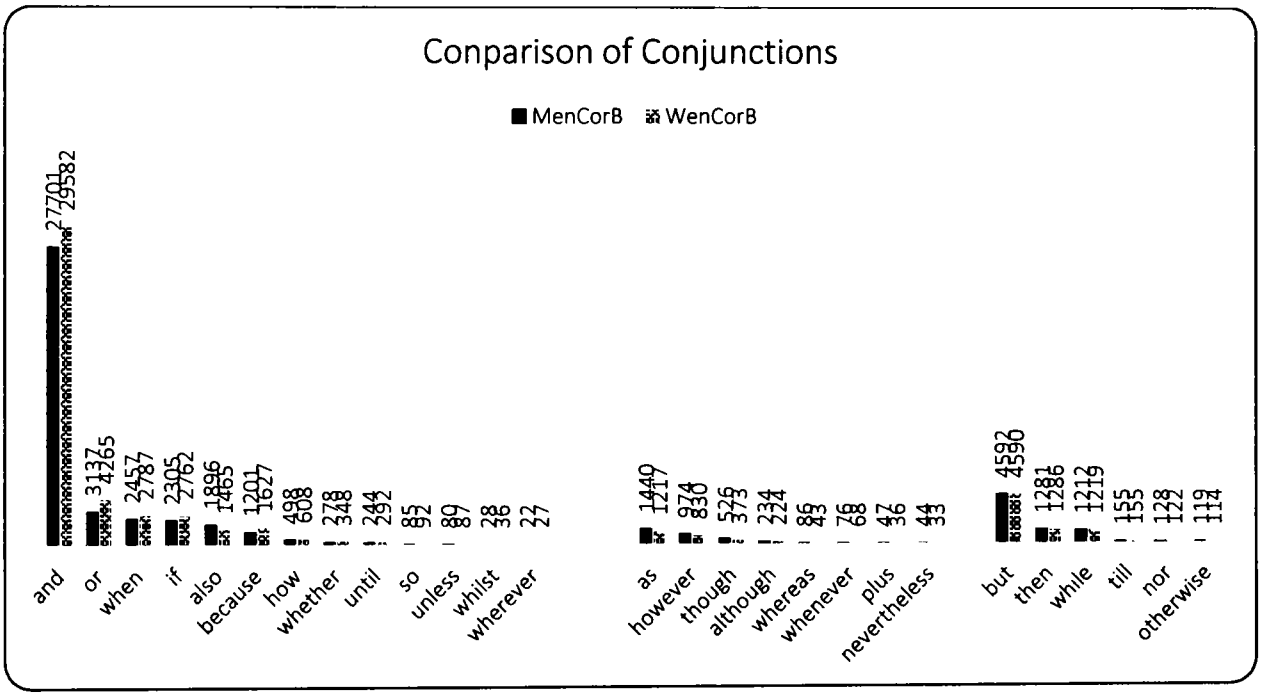


Figure 5.13. Conjunctions (frequencies normalised to one million)

GenCorB was further examined for gender differences in the use of specific conjunctions as given in Figure 5.13. It was observed than women used more *and*, *of*, *when*, *if*, *also*, *because*, *how*, *whether*, *until*, *so*, *unless*, *whilst*, and *wherever*. In contrast, men tended to overuse *as*, *however*, *though*, *although*, *whereas*, *plus*, and *nevertheless*. It was further observed that *but*, *then*, *while*, *till*, *nor*, and *otherwise* were used both by men and women at about the same frequency. The overall overuse of conjunctions by women bloggers is an indicator of their more complex style of writing in comparison to men bloggers. As Pennebaker (2013) argued, many of the conjunctions are also an indicator of a style where a speaker or writer makes a distinction, e.g. between what happened and what did not take place; what they thought and what they did not think about; or what included in a category and what not. He also pointed out that a complex thinker can make distinctions. In the light of this, it could be argued that women's overuse of conjunctions indicated their more complex thinking style.

5.5.3 Negations

Previous research on negation that could be consulted for discussion in this section consistently showed that women used negations far more frequently than men. For instance, Koppel, Argamon, and Shimoni (2002) analysed a corpus of 566 texts taken from the BNC to identify linguistic features more commonly used by one or the other gender. The data included both fictional and non-fictional texts. They asserted that the female indicators included negation as a language feature. Specifically, the study provided that women used *not* more than men for negation.

Later, Schler, Koppel, Argamon, and Argamon (2005) analysed a corpus of blogs consisting of about 300 million words for language differences on the basis of age and gender. They noted that for each age bracket, female bloggers used more negation words in comparison to men.

Similarly, Newman, Groom, Handelman, and Pennebaker (2008) compiled a large corpus of 14000 text samples and studied gender differences in language use. The results of their study also showed that women used more negations. In the same vein, in his book *The Secret Life of Pronouns*, Pennebaker (2013) includes negations in the list of language features that distinguish the style of men and women and declares it to be a feature of women's language. No research in the literature studied could be found, which contradicted this tendency in women's language.

The results of the present study were also consistent with the previous research. The Mann-Whitney U-test showed that women used significantly more negations in comparison to men (U = 843487; U (standardized) = -7.738; Expected value = 1014444.000; Variance (U) = 488099381.471; p-value (Two-tailed) = < 0.0001).

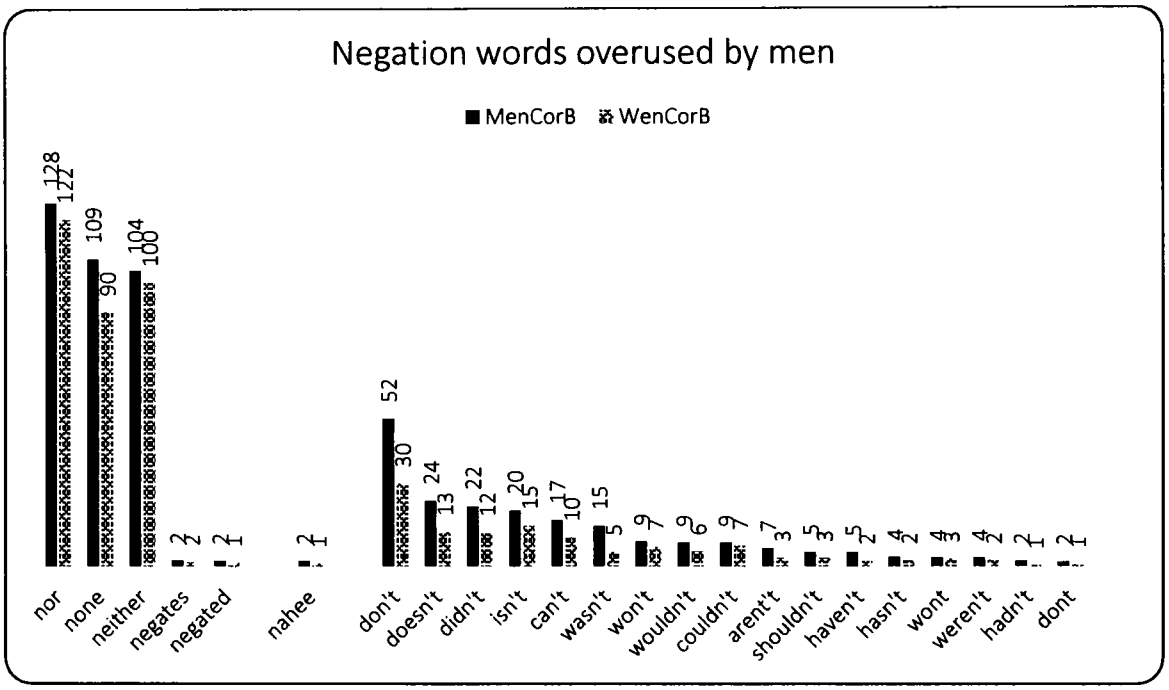


Figure 5.14. Negations used more by men (frequencies normalised to one million)

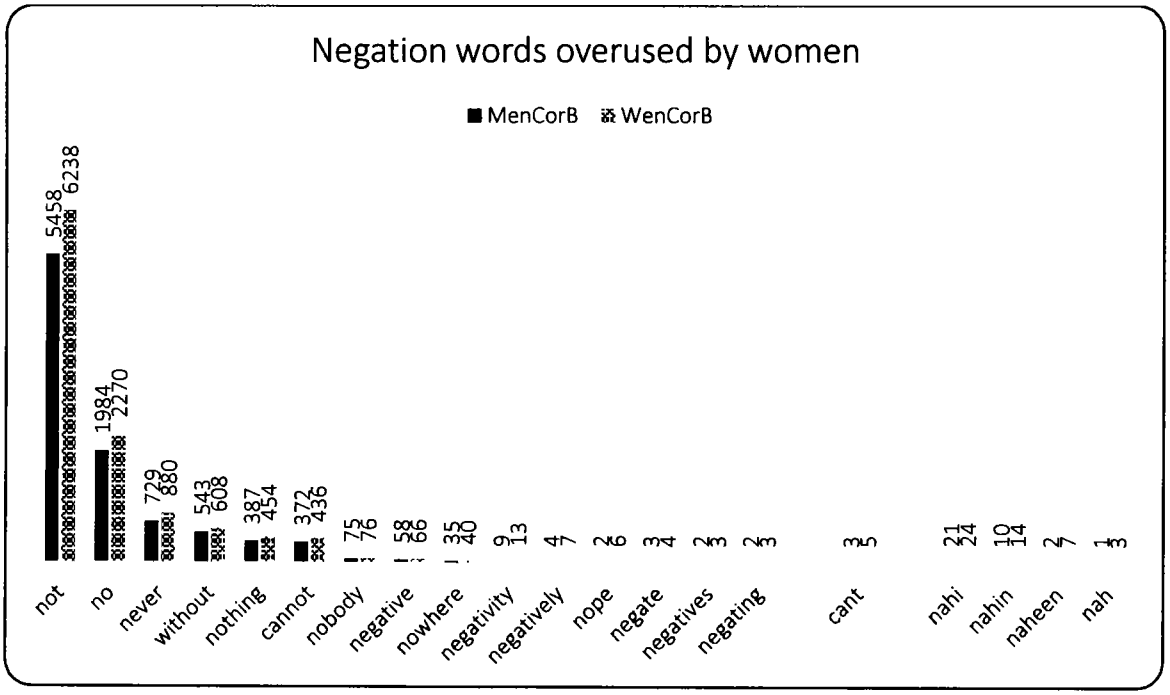


Figure 5.15. Negations used more by women (frequencies normalised to one million)

Following the hypothesis test for this language feature, both the gendered corpora were further explored for frequency differences between men and women bloggers in the use of particular negations, which were used with a raw frequency of 10 or more in either corpus. This analysis yielded interesting patterns. Besides having subtle nuances in the use of negations, both the genders were found to apply codemixing. Although LIWC2015 version used in this study has been based on English dictionary, the tool also surprisingly captured some of the Urdu expressions used in GenCorB for negation.

As shown in Figures 5.14 and 5.15, men used far more frequently the contracted forms of auxiliaries showing negations, except the expression *cant*, which was used more by women. On the other hand, women tended to overuse Urdu words *nahi*, *nahin*, *naheen* (i.e. 'no') and *nah* (i.e. 'not'), except *nahee* (i.e. 'no'), which is a spelling variation of *naheen* and not a separate expression. It is also interesting to note that both in English and Urdu, women used more *not* and *no* for negations. This observation is consistent with Koppel, Argamon, and Shimoni's (2002) result who studied gendered language features in the BNC and concluded that, in comparison to men, women used more *not* for negation.

Seen through a categorical-dynamic index (CDI), the overall more frequent use of negations has been considered as an indicator of dynamic language (personal narrative style) (Pennebaker, Chung, Frazee, Lavergne, & Beaver, 2014). So, it can be concluded that the nature of women bloggers' style tended to be that of a personal narrative.

5.5.4 Common Adjectives

The way people describe things exhibits male and female characteristics of language use, which is represented in a text with the use of adjectives. Previous studies have yielded

overwhelming evidence that women used more adjectives in comparison to men. However, some research reported contrasting results.

Probably it was Jespersen (1922) who first highlighted that women use certain adjective more than men. Later, Lakoff (1975) added that when making lexical choices (especially colours and adjectives), men tended to use categories at the basic or generic level, while women used categories at the subordinate level. For example, women discriminated between the shades of *beige*, *lavender*, *aquamarine*, which were absent from men's vocabulary. Women would use a different set of adjectives (*charming*, *divine*, *adorable*) than men (*cool*, *great*, *terrific*) to express their opinion on a subject. Similarly, Key (1975) summarised students' research work and concluded that women tended to use reduplicated forms such as '*teeny-tiny*, *itsy-bitsy*'. They also tended to use words that emphasized femininity, such as '*adorable*, *bubbly*, *cuddly*, *cute*, *darling*, *exquisite*, *pretty*, *precious*, and *sweet*'. They tended to use more emphatic forms of adjectives such as '*fantastic*, *horrifying*, *startling*'. Males tended to use forms that emphasized masculinity: '*barbed*, *bristly*, *leathery*, *lusty*'. In the same vein, Amir, Abidin, Darus, and Ismail (2012) studied Malaysian teenage bloggers who used blogs as a diary to express daily issues about life. Using a qualitative approach, they concluded that women used adjectives far more frequently than men. More recently, Hanafiyeh and Afghari (2014) examined whether men and women were different with respect to the use of certain expressions including empty adjectives. They used Lakoff's (1975) ideas concerning linguistic differences between males and females. Their results revealed that there were significant differences between the groups in the use of empty adjectives. The only study reported in this discussion that contradicted these results was of Argamon, Koppel, Fine, and Shimoni (2003), which reported that men used more adjectives.

The results of the present study contradicted some of the previous findings while it confirmed others. The Mann-Whitney U-test conducted for this study found that overall there was no significant difference in the use of common adjectives by men and women bloggers (U = 1069406.500; U (standardized) = 2.488; Expected value = 1014444.000; Variance (U) = 488109263.945; p-value (Two-tailed) = 0.013).

The data was further scrutinised to discover consistency in the use of specific adjectives identified by previous research. In this regard, findings pertaining to overuse of adjectives of colours based on Berlin and Kay's (1969) colour taxonomy and Key's (1975) list of adjectives were compared.

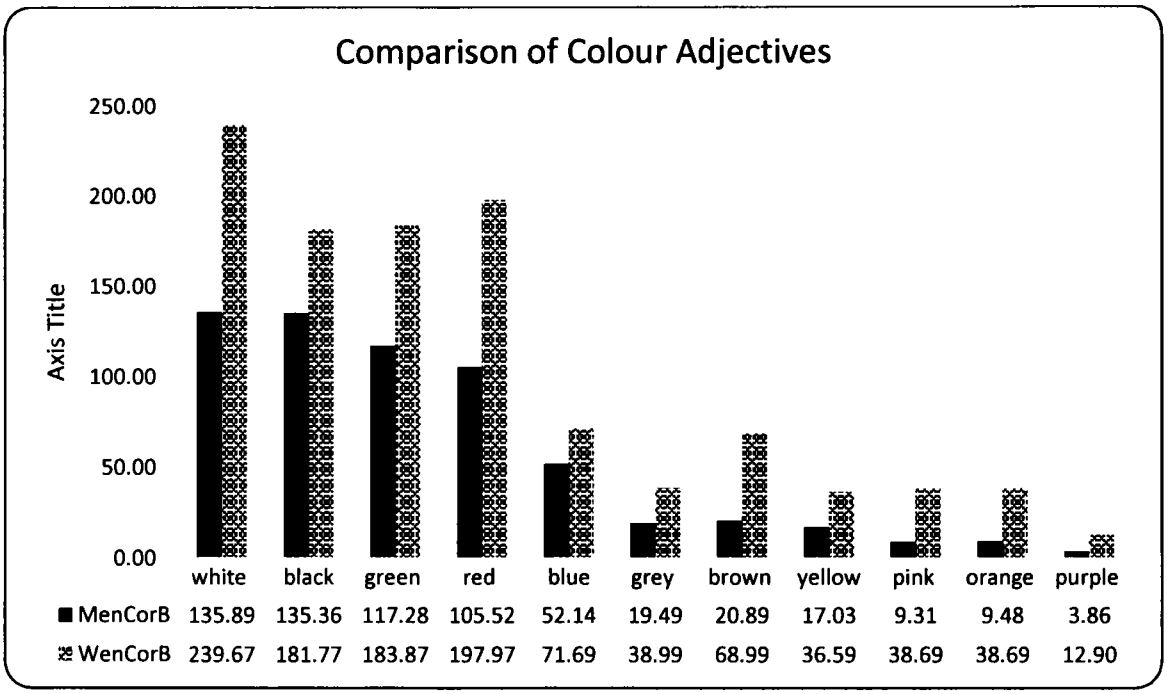


Figure 5.16. Colour adjectives based on Berlin and Kay's (1969) colour taxonomy (frequencies normalised to one million)

To find difference in adjectives of colours, Berlin and Kay's (1969) taxonomy was used for comparison who studied the colour terms that exist in a sample of the world's languages. From such data, they extracted a natural hierarchy of colour terms, which would enable one to predict

what other colour terms are or are not present in a given language. For example, a language that contains a term for red must also contain terms for *black* and *white*; a language that contains a word for *green* must also have one for *red*. Certain terms (e.g. *white* and *black*) do not have any obvious precedence between them and so are listed at the same point in the hierarchy. As shown in Figure 5.16, women were found to have overused colour terms more than men.

Table 5.4

Comparison of results for selected adjectives (frequencies normalised to one million)

Adjective	MenCorB N. Fr	WenCorB N. Fr	Findings	
			Key's	Present Study
pretty	88.31	143.08	Women	Consistent
sweet	32.48	90.29	Women	Consistent
precious	27.39	37.49	Women	Consistent
cute	10.01	20.10	Women	Consistent
horrifying	6.50	13.80	Women	Consistent
adorable	3.34	9.60	Women	Consistent
exquisite	6.67	8.10	Women	Consistent
cuddly	1.58	1.80	Women	Consistent
bubbly	1.05	1.50	Women	Consistent
fantastic	27.39	14.70	Women	Contradictory
darling	10.36	9.60	Women	Contradictory
startling	11.06	3.30	Women	Contradictory
barbed	2.63	3.90	Men	Contradictory
lusty	0.70	0.90	Men	Contradictory
leathery	0.00	0.60	Men	Contradictory
bristly	0.00	0.00	Men	Not found

Similarly, the adjectives identified by Key (1975), based on meta-analysis of students' research work, were also searched in GenCorB. As given in Table 5.4, mixed findings were recorded after comparison of present results with those of the previous research. While some of the findings were consistent, others were not.

5.5.5 Comparisons

In the literature accessed for discussion given in this section, no previous research could be found that focused on gender differences in the use of words that relate to comparisons. The present study, however, found that men used such words more in comparison to women. These findings, therefore, are important addition to knowledge about language features and gender differences. The Mann-Whitney U-test conducted for the present study found significant difference in the use of comparisons between the two genders (U = 1106456; U (standardized) = 4.165; Expected value = 1014444.000; Variance (U) = 488102029.761; p-value (Two-tailed) = < 0.0001). This tendency in men could be linked to their categorical and analytic thinking, which is the ability to clearly distinguish one thing from another.

5.5.6 Interrogatives

With regard to use of questions, previous research has produced contradictory results. Lakoff (1975) was probably the first to indicate that women use questions, particularly tag questions, more in comparison to men. Later, a number of studies (McMillan, Clifton, McGrath, & Gale, 1977; Mulac & Lundell, 1986) have reported greater female use of tag questions, although others (Dubois & Crouch, 1975) have found the opposite.

One of the earliest studies that examined gender differences in the use of questions was Fishman's (1980) who analysed fifty-two-minute recording of three US couples. The study recorded the frequency of questions asked by members of the two genders and observed that most of the questions were asked by women. The study ascribed this tendency in women to their effort to keep the conversation going. Two years later, Maltz and Borker's (1982) results were consistent

with Fishman's (1980) in that women considered questions as a tool to maintain conversation. In contrast, men viewed questions as a tool to seek information.

Although further studies (Coates, 1993) also reported consistent results that women used questions more frequently than men, they added another interpretative dimension, i.e. questions were used by women because their language was more tentative. Subsequent research (James & Clarke, 1993; Suborn, 2013) challenged this interpretation by going back to the previous position that women used more questions because they wanted to encourage others to speak, i.e. they used questions as a strategy for rapport-building. Some studies (Mulac, Wiemann, Widenmann, & Gibson, 1988) found that questions are more common in women's contributions to dyadic interactions, whereas directives that tell the audience to do something are more likely to be found in men's conversational contributions. To compare, Newman, Groom, Handelman, and Pennebaker (2008) did not find any gender differences in asking questions.

Some recent studies also did not replicate the findings that women used more questions. For instance, a study of e-mail communication (Thomson & Murachver, 2001) found that men and women were equally likely to ask questions; offer compliments, apologies, and opinions; and hurl insults at their "net pal." Others (Mulac, Seibold, & Farris, 2000) have reported significant differences in the opposite direction, i.e. men asked more questions.

The results of this study for interrogatives were consistent with the research, which found that women used more questions in comparison to men. The Mann-Whitney U-test conducted for the present research found that women used interrogatives more frequently than men (U = 819088; U (standardized) = -8.842; Expected value = 1014444.000; Variance (U) = 488098890.338; and p-value (Two-tailed) = < 0.0001).

5.5.7 Numbers and Quantifiers

Although numbers and quantifiers are two different variables in LIWC2015, they have been treated together in this section because they are somewhat related concepts as explained in this section. These two categories have not received much attention by the research studies in the past. Some of the previous research in CL (Koppel, Argamon, & Shimoni, 2002; Argamon, Koppel, Fine, & Shimoni, 2003; Herring & Paolillo, 2006; Newman, Groom, Handelman, & Pennebaker, 2008; and, Manjavacas, 2015) attributed both the categories to men's linguistic style. Others (Lenard, 2016), however, found no significant gender-based differences in the use of numbers and quantifiers.

The Mann-Whitney U-test conducted for the present study yielded consistent results for the variable of numbers but contradicted the previous research with regard to the variable of quantifiers. This test found that men used more numbers in comparison to women (U = 196724.500; U (standardized) = 8.251; Expected value = 1014444.000; Variance (U) = 488106490.648; p-value (Two-tailed) = < 0.0001). However, the test found no significant difference in the use of quantifiers in men and women bloggers (U = 966865.500; U (standardized) = -2.154; Expected value = 1014444.000; Variance (U) = 488100465.106; p-value (Two-tailed) = 0.031). The results of the present study for quantifiers were consistent with Lenard (2016), whose LIWC-based study on a corpus of political speeches found no significant difference in the use of quantifiers. The overuse of numbers and quantifiers has been linked by Pennebaker (2013) to formal and analytical thinking respectively. Therefore, the use of the two language features in MenCorB could be an indicator of men's formal and analytical thinking. The higher score of men on LIWC2015 *analytical thinking* category in this study also supports this tendency.

5.6 Psychological Processes

5.6.1 Affective Processes

Perhaps a persistent idea about gender differences is that women tend to be more emotional than men. Previous research on stereotypes has rated emotionality as an important dimension of gender differences: women have been reported to be more sensitive to others' emotions, more expressive, excitable and easily hurt in comparison to men (Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972; Ruble, 1983; Ashmore, Del Boca, & Wohlers, 1986; Lutz, 1990; Fabes & Martiṅ, 1991). A possible interpretation for these stereotypes is based on Eagly's *Social Role Theory*, which argues that women and men differ in their social roles (Aries, 1996; Bales, 1950; Eagly, 1987). Instrumental roles, traditionally more often held by men, have more status and are associated with competence, rationality, and autonomy, whereas socio-emotional roles, which are more often occupied by women, are based on their caring role within the family context. Because men more often play instrumental roles and women more often socio-emotional roles, men and women are also assumed to be different, women being more warm, emotional, and submissive while men are more rational, dominant, and decisive (Fischer, 2011).

Numerous research projects conducted from sociological and psychological perspectives have encouraged linguists to study differences between men and women in expressing emotions. According to Trudgill (2000), differences between men and women in expressing emotions in language are behaviour patterns caused by different social roles. Although all human beings express emotions, they do it through different strategies (Cameron, 2005).

Much literature in the past has been produced on analysis of gender differences in expressing emotions but with conflicting results. For instance, studies by Gilligan (1982), Mulac, Studley, and Blau (1990), Jordan, Surrey, and Kaplan (1991), Fivush and Buckner (2000), and

Thomson and Murachver (2001) demonstrated that women express more emotions while men restrained from expressing their feelings. A recent study on gender and emotion differences in online collaboration by Iosub, Laniado, Castillo, Fuster Morell, and Kaltenbrunner (2014) pointed to status as an important variable in expressing emotions. They found that female regular editors were the most relationship-oriented promoting social affiliation and emotional connections, while male administrators were the least emotional. To paraphrase, the higher status a person had, the less emotional they were. Similarly, analysing political speeches of male and female congress members, Lenard's (2016) results were also consistent with the research, which had found women as overusing emotional language in comparison to men. In contrast, in a study on managers criticisms, Mulac, Seibold, and Farris (2000) found that male managers were more emotional than female thus not confirming to the stereotypes.

The results of the present study also supported the idea that women were more emotional in comparison to men. The Mann-Whitney U-test conducted for the present study found a statistically significant gender difference, with women overusing the category (U = 904212.500; U (standardized) = -4.989; Expected value = 1014444.000; Variance (U) = 488112689.323; p-value (Two-tailed) = < 0.0001).

Positive and Negative Emotions

The study of affective process can further be extended to studying gender differences in expressing positive and negative emotions separately. LIWC2015 treats positive and negative emotions as separate variables. However, for convenience in discussion, the two variables are treated together in this section.

Previous research on gender differences in expression of positive emotions has yielded consistent results. For instance, Iosub, Laniado, Castillo, Fuster Morell, and Kaltenbrunner (2014) found that women expressed these emotions with far greater frequency than men in online context. Their results supported those of Mehl and Pennebaker (2003), Newman, Groom, Handelman, and Pennebaker (2008), Kivran-Swaine, Brody, Diakopoulos, and Naaman (2012), and Schwartz, et al. (2013). In other words, women were reported, in both online and offline contexts, to have expressed positive feelings more than men.

While results for by the previous research for positive emotions were consistent, those for negative emotions were contradictory. For instance, Mehl and Pennebaker (2003), and Newman, Groom, Handelman, and Pennebaker (2008) observed that men used more negative emotions in comparison to women whereas Mulac, Studley, and Blau (1990), Thomson and Murachver (2001) found the opposite, i.e. women used more negative emotions. Finally, even though female's emotion words were more active, Fischer (1993), and Bell, McCarthy, and McNamara (2006) did not find a statistically significant gender difference.

The positive emotions mean ranks (M = 2.93 for men and M = 2.98 for women) indicated that the women indeed expressed positive feelings more than the men; however, the Mann-hitney test results showed that the difference was not statistically significant (U = 994651.500; U = (standardized) -0.896; Expected value = 1014444.000; Variance (U) = 488110082.580; p-value (Two-tailed) = 0.370). These results support those of Fischer's (1995), and Bell, McCarthy, and McNamara (2006).

Table 5.5

Comparison of top 20 positive emotion words (frequencies normalised to one million)

S No	MenCorB			WenCorB		
	Word	Raw Fr	N. Fr	Word	Raw Fr	N. Fr
1	well	6289	1104.16	well	3255	976.36
2	good	4363	766.01	good	2786	835.68
3	party	3950	693.50	love	2309	692.60
4	best	3320	582.89	better	1739	521.62
5	better	2803	492.12	best	1634	490.13
6	great	2681	470.70	great	1233	369.84
7	play	2402	421.72	important	1128	338.35
8	love	2290	402.06	sure	1108	332.35
9	won	1985	348.51	hope	1050	314.95
10	support	1930	338.85	support	1004	301.15
11	played	1899	333.41	free	988	296.35
12	security	1818	319.19	kind	960	287.95
13	parties	1795	315.15	play	835	250.46
14	win	1708	299.87	security	835	250.46
15	playing	1700	298.47	care	767	230.06
16	important	1687	296.19	happy	754	226.16
17	free	1632	286.53	won	748	224.36
18	hope	1628	285.83	certain	738	221.36
19	sure	1612	283.02	beautiful	713	213.87
20	peace	1321	231.93	peace	706	211.77

Since a comparison of all the emotion words captured by LIWC2015 was not possible, the top twenty words used by either gender to express positive emotions in both the corpora were selected and compared as shown in Table 5.5. According to the results, we can draw a conclusion that even though men and women used positive feelings at the same frequency, men focused more on words related to playing (*play*, *played*, *playing*) whereas women have used expressions like *care*, *kind*, *happy* and *beautiful*. Similarly, other expressions like *won* by men and *love* by women have been used with different frequency and have quite distant positions in the order of words in both the lists. This tendency indicates the gender preference for expressions of positive feelings,

i.e. women prefer words that indicate feelings of love and care while men prefer words that indicate sense of achievement.

In contrast to positive emotions, the results of Man-Whitney U-test conducted for this study pointed to a statistically significant difference on negative emotions use (U = 894530; U (standardized) = -5.428; Expected value = 1014444.000; Variance (U) = 488110507.920; p-value (Two-tailed) = < 0.0001). Women were found to have used more negative emotions in comparison to men. These results supported findings reported by Mulac, Studley, and Blau (1990), Gross and John (1998), Seidlitz and Diener (1998), Thomson and Murachver (2001), O'Kearney and Dadds (2004), and, Chraif and Anitei (2013).

Table 5.6

Comparison of Top 20 negative emotion words (frequencies normalised to one million)

S No	MenCorB		WenCorB	
1	war	459.29	violence	370.75
2	lost	356.93	war	317.36
3	attack	297.42	lost	315.26
4	wrong	252.12	wrong	309.26
5	problem	250.36	bad	272.36
6	poor	240.71	rape	249.87
7	bad	235.79	killed	247.47
8	failed	215.42	problem	244.17
9	violence	214.37	fear	239.07
10	killed	213.32	poor	236.97
11	attacks	201.55	attack	218.97
12	problems	196.99	alone	203.37
13	terrorism	188.56	problems	195.27
14	unfortunately	174.87	victims	195.27
15	fight	169.07	terrorism	191.07
16	fear	156.08	attacks	181.47
17	alone	148.36	abuse	180.27
18	serious	144.67	hate	172.78
19	difficult	144.49	pain	161.08
20	terrorists	143.44	killing	158.38

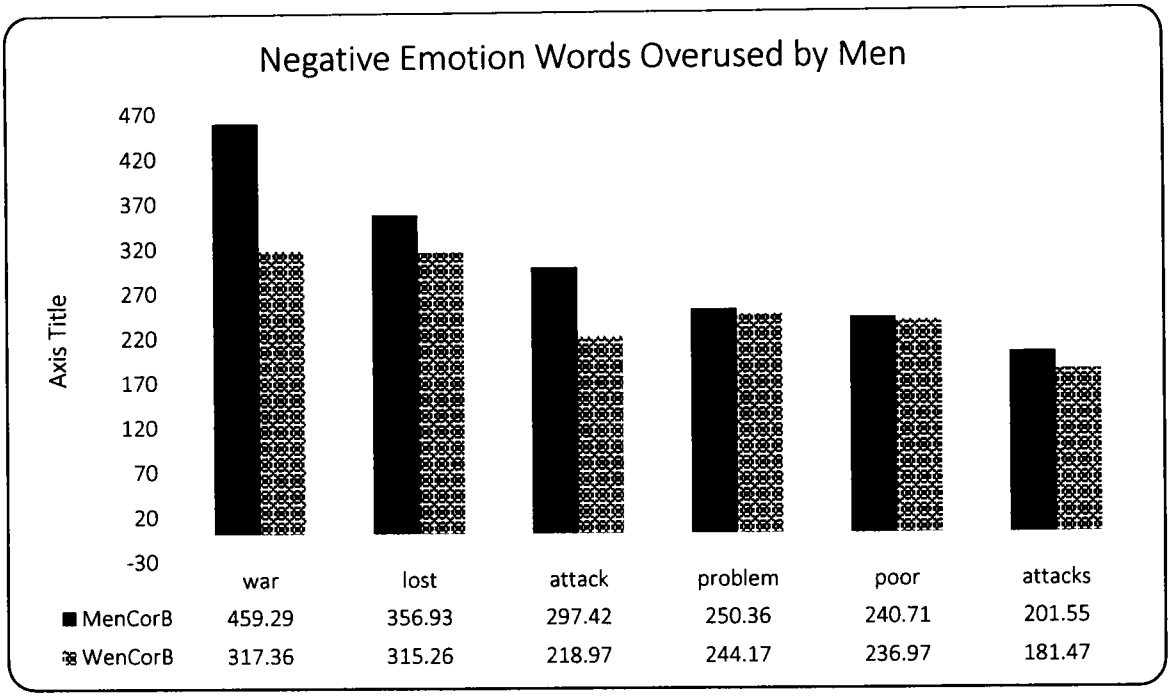


Figure 5.17. Negative emotion words used more by men (frequencies normalised to one million)

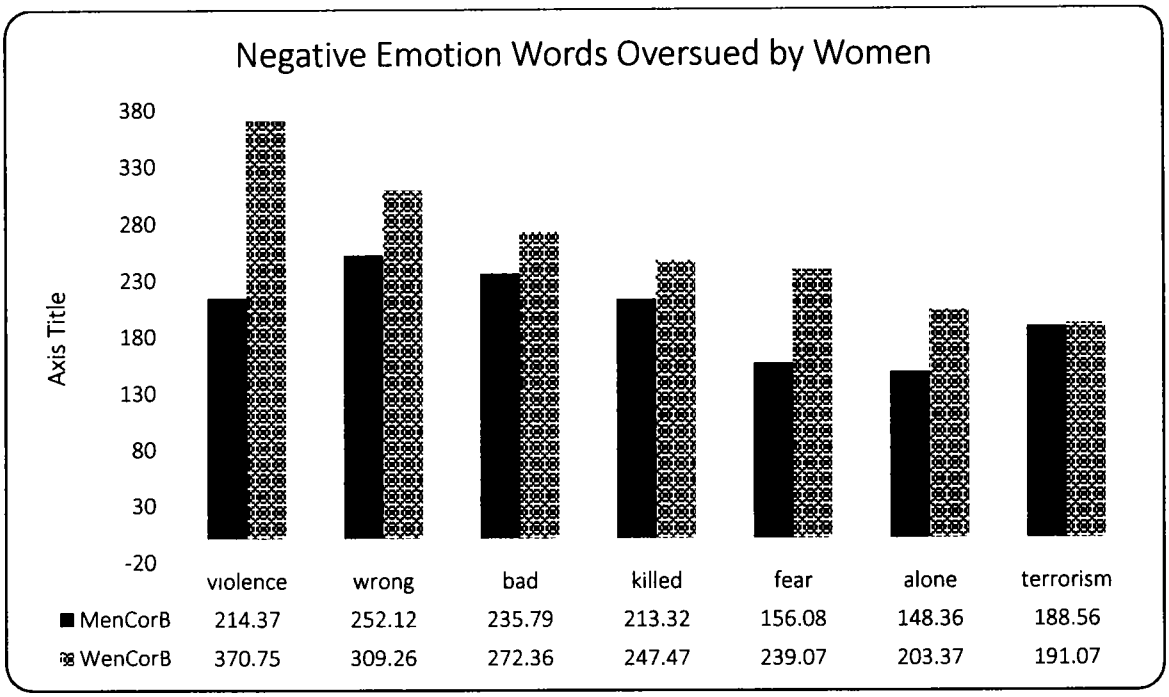


Figure 5.18. Negative emotion words used more by women (frequencies normalised to one million)

In addition, top 20 words showing negative emotions were selected from each of the corpora as shown in Table 5.6. It was observed that men used more words like *war*, *lost*, *attack* and *attacks* whereas women used more words like *wrong*, *fear*, *alone*, and *killed* as shown in Figures 5.17 and 5.18. Additionally, words like *rape*, *victims*, *abuse*, *hate*, *pain*, and *killing* are used by women only, which show their concerns about victimisation in general and women's issues in particular. Here also, men tend to overuse expressions that relate to competition (*war*, *attack*) whereas women overuse words that relate to hate and victimisation.

Anxiety

Although primarily a domain of psychology, the study of realising anxiety in language has recently received attention from linguistics. In previous research, Mulac, Studley, and Blau (1990), Thomas and Murachver (2001), (Yu, 2014), and Manjavacas (2015) found that women express anxiety in their language more in comparison to men. Contrary to these findings, some studies (e.g. Xu, Li, Stefanone, & Fu, 2014; Lenard, 2016) yielded contrasting results and reported that men and women did not use anxiety words with different frequencies.

The results obtained from the Mann-Whitney U-test conducted for this study supported research, which found that women tend to use anxiety words far more frequently than men. The test found significant difference between genders (U = 876509; U (standardized) = -6.246; Expected value = 1014444.000; Variance (U) = 487750637.386; p-value (Two-tailed) = < 0.0001).

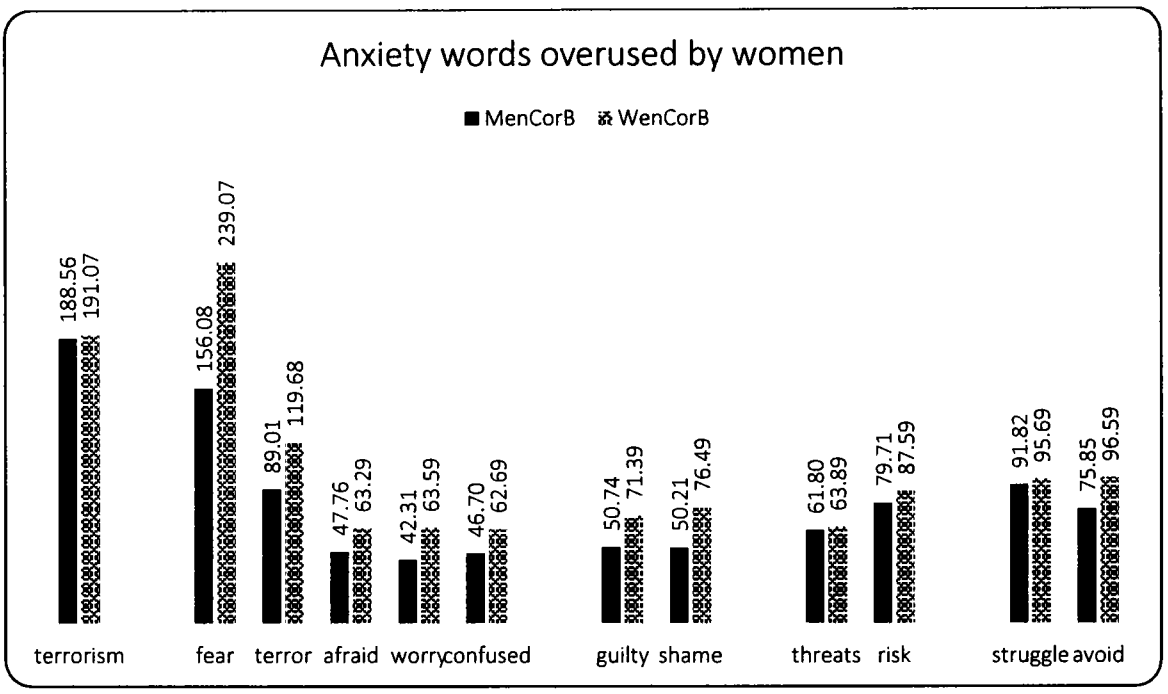


Figure 5.19. Anxiety words used more by women (frequencies normalised to one million)

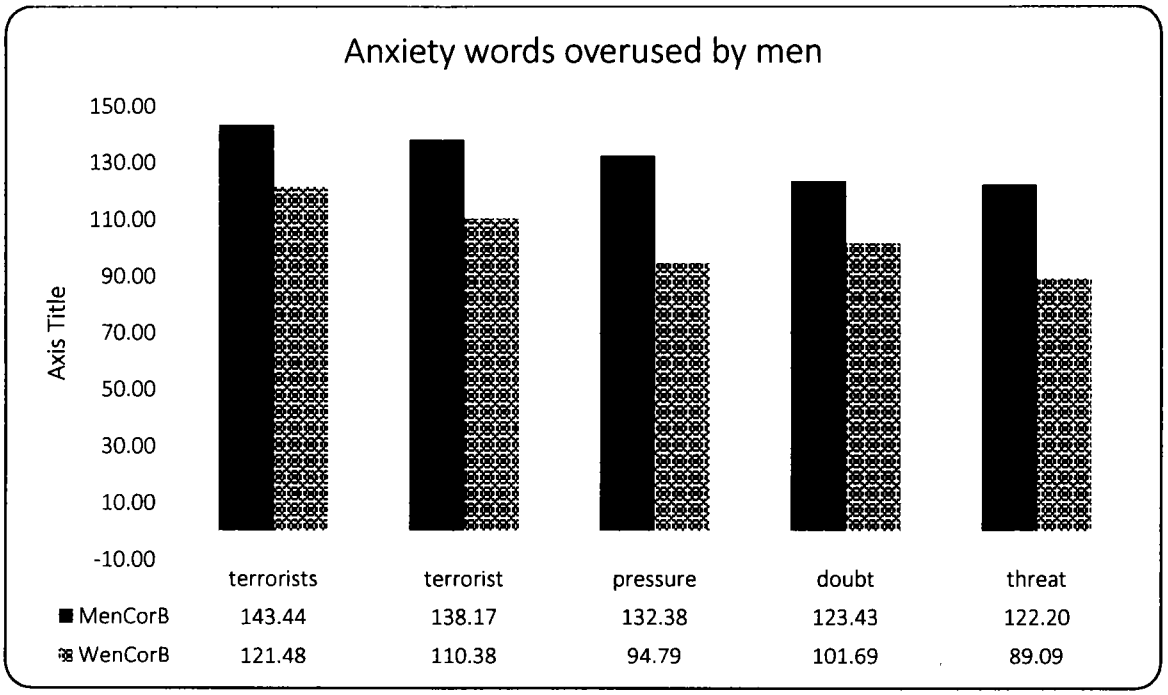


Figure 5.20. Anxiety words used more by men (frequencies normalised to one million)

Additionally, the data was analysed for specific anxiety words used more by men and women. For this purpose, top 20 frequent words from each corpus were studied. As Figures 5.19 and 5.20 show women used more terms that related to internal states of anxiety, e.g. *fear, guilt, shame.* In contrast, men used anxiety terms related to external causes of anxiety, e.g. *pressure, terrorist, threat.*

Anger

The study of anger is a central area in research related to language of emotions. Some research (Rudman, 1998; Heilman, 2001) related anger to the language of men and argued that society does not expect women to express anger. Therefore, women are discouraged to show anger in their language. Lakoff (2004) gave another explanation for lesser representation of anger in women's language in comparison to men. She claimed that while women expressed anger, the power of anger was denied to them: "You're so cute when you're mad" (p.22). She added that this kind of situation intensified male power and female powerlessness.

Previous research on anger expressed in language yielded contradictory results. Indeed, some researchers (Plant, Hyde, Keltner, & Devine, 2000; Mehl & Pennebaker, 2003; Wang & Hsieh, 2007; and Schwartz, et al., 2013) confirmed that men expressed their emotions by using anger-related references more than women. On the other hand, studying the effect of gender on emotions in European Americans and Hmong Americans, Chentsova-Dutton and Tsai (2007) not only reported that in comparison to men women in both groups experienced more intensity in expressing emotions but also observed that the gender effect on emotions held across groups from diversified ethnic backgrounds. In the same vein, another study (Jones, 2016) examined the hypothesised change in language of Hillary Clinton in her political career over the years and

found that she tended to use a masculine index of anger the most when she became Secretary of State (2009-2013), which confirmed the study's presumption that her language had become more masculine over the passage of time. The studies that did not replicate the findings of gender differences in expressing anger were Newman, Groom, Handelman, and Pennebaker (2008) and Lenard (2016).

The results of Mann-Whitney U-test conducted for this study found no significant difference in the language of men and women bloggers in the use of words related to anger (U = 988044.500; U (standardized) = -1.195; Expected value = 1014444.000; Variance (U) = 487972585.565; p-value (Two-tailed) = 0.232). These results were in line with those of Newman, Groom, Handelman, and Pennebaker (2008), and Lenard (2016).

Sadness

Lakoff (2004) reported that traditionally women were perceived to express sadness and cry more than men, which is considered as an expression of women's powerlessness. However, research on expression of sadness by gender has produced conflicting results. Mulac, Studley, and Blau (1990), Thomas and Murachver (2001), Baumann, Krasnova, Veltri, Yusni (2015), and Lenard (2016) reported that women expressed sadness more than men. However, several researchers examined the stereotype and found contrary results. Wang and Hsieh (2007) reported in their study that boys were more expressive of sadness in comparison to girls. They attributed this tendency in girls to their focus on maintaining harmonious relationships.

The results of Mann-Whitney U-test for this study found significant difference in the use of expressions of sadness by men and women. The study found that women used more words that show sadness in comparison to men (U = 900055.500; U (standardized) = -5.178; Expected value

= 1014444.000; Variance (U) = 487941861.453; p-value (Two-tailed) = < 0.0001). These results support the finding of Mulac, Studley, and Blau (1990), Thomas and Murachver (2001), Baumann, Krasnova, Veltri, Yusni (2015), and Lenard (2016). A further analysis of the top 20 most frequent expressions of sadness from each corpus was made. It was found that in these lists the words *tears*, *cry* and *depression* were used by women but not by men. Further analysis of the overall use of these three expressions in both the corpora also confirmed that women used them more in comparison to men.

5.6.2 Social Processes

This category of LIWC2015 contains all expressions that relate to human relations in general, both male and female, friends and family. Traditionally, women have been considered as more social in comparison to men (Lenard, 2016). Several studies in the past have confirmed this trend in women. For example, a study conducted by Brownlow, Rosamond, and Parker (2003) showed that women expressed social processes more in their language. In CL research, a study by Newman, Groom, Handelman, and Pennebaker (2008) also confirmed such trends in the language of female participants, especially their far greater reference to family relations. In contrast, Bell, McCarthy, and McNamara (2006) found no significant gender differences in use of words that referred to social processes.

The Mann-Whitney U-test conducted for the present study found significant difference between genders in the use of words related to social processes (U = 684032.500; U (standardized) = -14.955; Expected value = 1014444.000; Variance (U) = 488115745.652; p-value (Two-tailed) = < 0.0001). It found that women used more words related to social processes in comparison to

men. These results confirmed the findings reported by Brownlow, Rosamond, and Parker (2003), and Newman, Groom, Handelman, and Pennebaker (2008).

Family, Friends, Female/Male References

LIWC2015 calculates percentages separately for words related to the linguistic variable of family, friends, female and male references. However, for the sake of convenience in discussion, the four variables have been treated together in this section.

Some of the previous research (Schler, Koppel, Argamon, & Pennebaker, 2005; Newman, Groom, Handelman, & Pennebaker, 2008) has supported the notion that female's language is more oriented towards others (i.e. towards family members, friends and humans). However, such findings have not been replicated by others. For instance, more recently, Xu, Li, Stefanone, and Fu (2014) conducted a LIWC-based study on news commenting behaviour by social media users and reported no evidence that women used references to others far more frequently than men.

The Mann-Whitney U-test conducted for the present research found significant difference between the two genders for words related to family. The results showed that women used more family words in comparison to men (U = 668868; U (standardized) = -15.778; Expected value = 1014444.000; Variance (U) = 479689264.338; p-value (Two-tailed) = < 0.0001).

After obtaining the overall hypothesis test for this language variable, the data was further scrutinised for specific words related to the category of family. It was found that almost all the words related to family were used more by women. However, the words pertaining to the theme of 'marriage' were specially used more frequently by women than by men as shown in Figure 5.21.

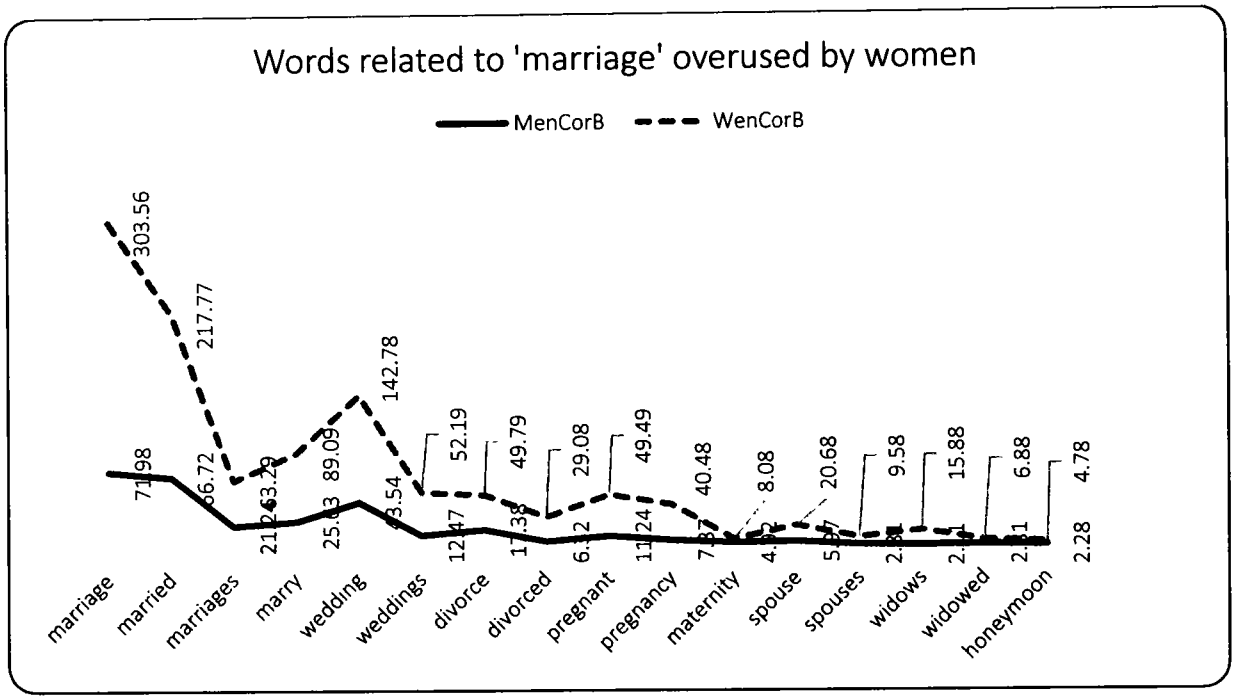


Figure 5.21. Words related to the theme of 'marriage' used more by women (frequencies normalised to one million)

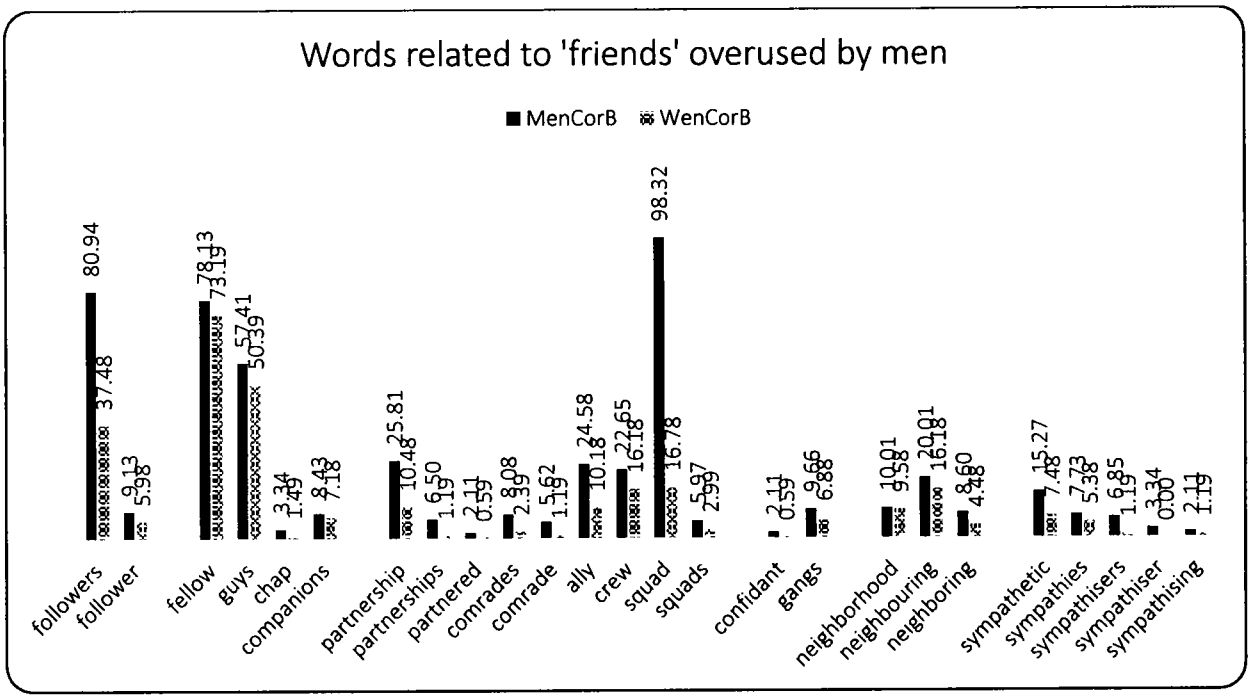


Figure 5.22. Words related to 'friends' used more by men (frequencies normalised to one million)

Similarly, Mann-Whitney U-test conducted for words related to the variable 'friends' also showed a significant gender difference. It was found the women used expressions related to friends and friendship more than men (U = 868430.500; U (standardized) = -6.654; Expected value = 1014444.000; Variance (U) = 481532249.829; p-value (Two-tailed) = < 0.0001).

The data was further examined to find out more subtle differences in the use of specific expressions by men and women. It was found that women used words, which indicated friendship based on personal association and love. In contrast, men used more words, as shown in Figure 5.22, that pertained to leader-follower relationship (*follower*, *followers*), ordinary colleagues (*fellow*, *guys*), business relationship (*partner*, *ally*), relationship based on common cause (*gang*, *confidant*), or relationship on humanitarian grounds (*sympathy*).

The Mann-Whitney U-test conducted for words referring to female gender also showed significant gender differences. The results indicated that women use words referring to female gender far more frequently than men (U = 564504; U (standardized) =-20.562; Expected value = 1014444.000; Variance (U) = 478829416.710; p-value (Two-tailed) = < 0.0001). However, the hypothesis test conducted for use of words related to male gender found no difference between women and men (U = 1073662; U (standardized) = 2.681; Expected value = 1014444.000; Variance (U) = 487735056.629; p-value (Two-tailed) = 0.007). This tendency further supports the finding of gender polarity between men and women as indicated by the present study during discussion on some previous language features.

5.6.3 Cognitive Processes

Cognitive processes relate to the thinking models of human beings. LIWC2015 category of Cognitive Processes encompasses the sub-categories of thinking related to *insight*, *causation*, *discrepancy*, *tentativeness*, *certainty*, and *differentiation*.

Previous research on analysis of gender differences in *cognitive processes* has given conflicting results. A study on spoken language by Poole (1979) found that women used more verbs related to thinking processes. Similar results were found by Mulac and Lundell (1994) who analysed written texts. Pennebaker (2013) also found that women used more cognitive words than men and considered it contradictory to the claim of Aristotle who believed men to be thinking more and women incapable of philosophical thought. More recently, Pennebaker, Boyd, Jordan, and Blackburn (2015) made a comparison between Hillary Clinton and Bernie Sanders and reported that Hillary Clinton used more words related to cognitive processes than Bernie Sanders.

In contrast to the above, Graells-Garrido, Lalmas, and Menczer (2015) conducted a study on people's biographies on Wikipedia with 84% male contributors. They found that the description of men contained words, which related to cognitive processes and work concerns as these two aspects are generally considered to have more significance in the lives of men. However, Kapidžić and Herring (2011) studied online conversation in chat messages of teenagers but found no evidence of gender differences. Similarly, Lenard (2016) analysed political speeches by congressmen and congresswomen but found no significant difference between the two genders in the use of words related to cognitive processes.

The Mann-Whitney U-test conducted for this study found a significant difference between men and women in the use of words related to cognitive processes. The results showed that women used more these words in comparison to men (U = 856381.500; U (standardized) = -7.154; Expected value = 1014444.000; Variance (U) = 488114977.092; p-value (Two-tailed) = < 0.0001).

Insight

As a sub-category of words showing *cognitive processes*, words related to *insight* are those that express self-reflection (Pennebaker, 2013). This category of words shows how much a

writer or speaker is self-referent and focused on the inner meaning of a subject. Previous research on analysis of gender differences based on this particular language dimension has produced mixed results. A study conducted in Australian context by Lester (2004), showed that the suicide notes written by women contained more insight words, which indicated that they were more self-referent. Similarly, the analysis of 14,000 text samples by Newman, Groom, Handelman, & Pennebaker (2008) also found that women used insight words more than men. However, Epstein, Sloan, and Marx (2005) who examined gender differences in psychological and physical symptom changes associated with written disclosure found that men used more words related to insight.

The Mann-Whitney U-test conducted for this study found significant gender differences in use of word that related to insight. The results showed that women used such words far more frequently than men (U = 850199; U (standardized) = -7.434; Expected value = 1014444.000; Variance (U) = 488104994.710; p-value (Two-tailed) = < 0.0001). Thus, the results indicated that women bloggers used self-reflection more than men bloggers.

Causation

As early as in 1980s, Maltz and Borker (1982), and Gilligan (1982) opined that since women were more oriented towards others and nurturing, they made more use of causal words in comparison to men. This view, however, was challenged by some of the subsequent studies. For instance, Epstein, Sloan, and Marx (2005), and Graells-Garrido, Lalmas, and Menczer (2015) reported contradictory results. They reported men's greater use of causality words. They interpreted these findings in terms of men's tendency to be more oriented towards problems. Amidst these two poles of opinions formulated by previous research, the studies of Newman,

Groom, Handelman, and Pennebaker (2008), and Manjavacas (2015) did not report any significant gender difference in the use of words related to causation.

The Mann-Whitney U-test conducted for the present study also did not find any significant gender difference in the use of words related to this language variable (U = 979446.500; U (standardized) = -1.584; Expected value = 1014444.000; Variance (U) = 488100859.133; p-value (Two-tailed) = 0.113). These results supported the findings of Newman, Groom, Handelman, and Pennebaker (2008), and Manjavacas (2015) for this variable.

Discrepancy

The use of discrepancy words shows a speaker's dissatisfaction or his/her desire for changes. This category of words is made up of modal verbs, which occur in the company of words that show some desire, hopes, regrets or ideals. This LIWC2015 category suggests discrepancy between how the world is and how it should, could, must, or ought to be (Lenard, 2016). Previous research by McMillan, Clifton, McGrath and Gale (1977), Biber, Conrad, and Reppen (1998), Mulac, Seibold, and Farris (2000), Mehl and Pennebaker (2003), and Newman, Groom, Handelman, and Pennebaker (2008) found that women use discrepancy words, especially the modal verb *could*, more in their language. They ascribed this tendency in women's language to their tentative style. However, Lenard (2016) reported no gender-based differences in the use of discrepancy words between men and women.

The Mann-Whitney U-test conducted for the present study found significant gender difference in the use of words related to this category. The result showed that women used these words far more frequently than men (U = 930409; U (standardized) = -3.804; Expected value =

1014444.000; Variance (U) = 488098884.246; p-value (Two-tailed) = 0.000). This trend in women's language is an indicator of their tentative language.

Tentativeness

Although much earlier Zimmerman and West (1975) had noted a connection between the language of females and tentativeness, it was Lakoff (1975) who concluded that women's language was tentative, and, thus, powerless, since they used more frequent hedges, tag questions, intensifiers and disclaimers/qualifiers in comparison to men. Her seminal work triggered debates and drew much attention to the aspect of tentativeness in women's language in diversified fields including feminist studies (e.g. Crawford, 1995), psychology (e.g. Leaper & Ayres, 2007) and linguistics (e.g. Tannen, 1994).

In linguistics, while some of the research (e.g. Carli, 1990; McConnell-Ginet, 2011) agreed with Lakoff (1975) and viewed tentativeness in women's language as a sign of their uncertainty, insecurity, and incompetence, making their language to be taken less seriously than that of men, others (e.g. O' Barr and Atkins, 1998) challenged Lakoff's (1975) argument on the basis that tentative language was not necessarily a sign of powerless language since such style was also used by people who enjoy power positions and those who are uneducated. Others (e.g. Harris, 1984) based their argument on the use of tag questions by powerful persons and interpreted its usage not as a sign of uncertainty but as a request for confirmation. Yet some others (e.g. Mulac, Giles, Bradac, & Palomares, 2013; Watts, 2003) linked tentativeness in women's language to their politeness and their intent to affiliate with or accommodate, facilitate and include others in communication.

Some of the research also reported that, besides gender, other factors, such as sex-based composition of communication groups, also paly important role in determining stylistic variations. For instance, Reid, Keerie, and Palomares (2003), Palomares (2008), and Palomares (2009) studied tentativeness in language on the basis of difference in topics and sex-group composition. They reported that both female and males were more tentative with masculine and feminine topics respectively in intergroup communication as compared to their linguistic behaviour in intragroup communication. Not all research, however, has reported tentativeness as a gender marker in communication. For instance, without focusing on topic-based contexts, some studies (e.g. Crosby & Nyquist, 1977; Schmader, Whitehead, & Wysocki, 2007; Newman, Groom, Handelman, & Pennebaker, 2008) did not report that men and women differed on the basis of tentativeness in language style.

The Mann-Whitney U-test conducted for the present study found significant difference between men and women in the use of tentative language. The results showed that women used more tentativeness in comparison to men (U = 890362.500; U (standardized) = -5.616; Expected value = 1014444.000; Variance (U) = 488107365.695; p-value (Two-tailed) = < 0.0001).

Certainty

This language variable of LIWC2015 dictionary consists of intensive adverbs, adverbials of frequency, and some modal auxiliaries (e.g. *must*). All of these language dimensions have received some attention by previous research. Much of the previous research on both writing and speaking genres of language accessed for discussion here (i.e. McMillan, Clifton, McGrath, & Gale, 1977; Mulac & Lundell, 1986; Biber, Conrad, & Reppen, 1998; Jaffe, Lee, Huang, & Oshagan, 1999; Mulac, Seibold, & Farris, 2000; Mondorf, 2002; Mehl & Pennebaker, 2003;

Pennebaker, Mehl, & Niederhoffer, 2003; and, Newman, Groom, Handelman, & Pennebaker, 2008) has reported that women have consistently made use of certainty words with far greater frequency when compared to men. Such research was based on the belief that certainty words are typical features of female linguistic style. Some studies, however, did not report any statistically significant difference between genders on words expressing certainty (Schmader, Whitehead, & Wysocki, 2007; Lenard, 2016).

The Mann-Whiteny U-test conducted for the present study found statically significant difference between genders with respect to this variable. The results showed that women used more expressions related to certainty with far greater frequency than men (U = 904557.500; U (standardized) = -4.974; Expected value = 1014444.000; Variance (U) = 488095748.844; p-value (Two-tailed) = < 0.0001). So these results were in line with the previous research quoted above, which found overuse of certainty expressions as a typical feature of women's language.

Although the overall results for this study showed that women tended to overuse certainty expressions in general, a further probe was conducted to see which expressions were particularly used more by men and women. A normalised frequency analysis of the top 100 most frequent expressions was carried out for this purpose. This analysis indicated a pattern of words that was used by the two genders. As shown in Figures 5.23 and 5.24, women tended use more words related to generalisation, conclusiveness or entirety, e.g. *all*, *everything*, and *totally*. In contrast men tended to use more words that referred to specificity or exactness, e.g. *especially*, *exactly*, and *namely*. This tendency in genders could be regarded as another indicator of stronger analytical thinking of men in comparison to women.

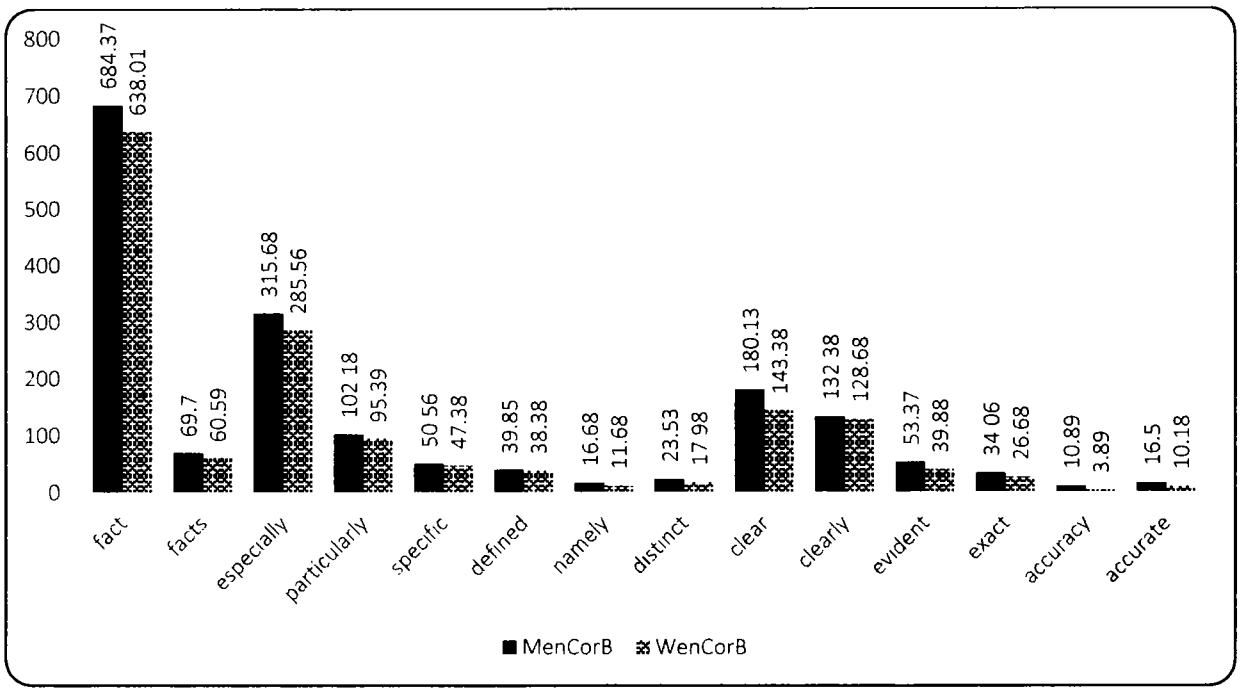


Figure 5.23. Certainty words used more by men (frequencies normalised to one million)

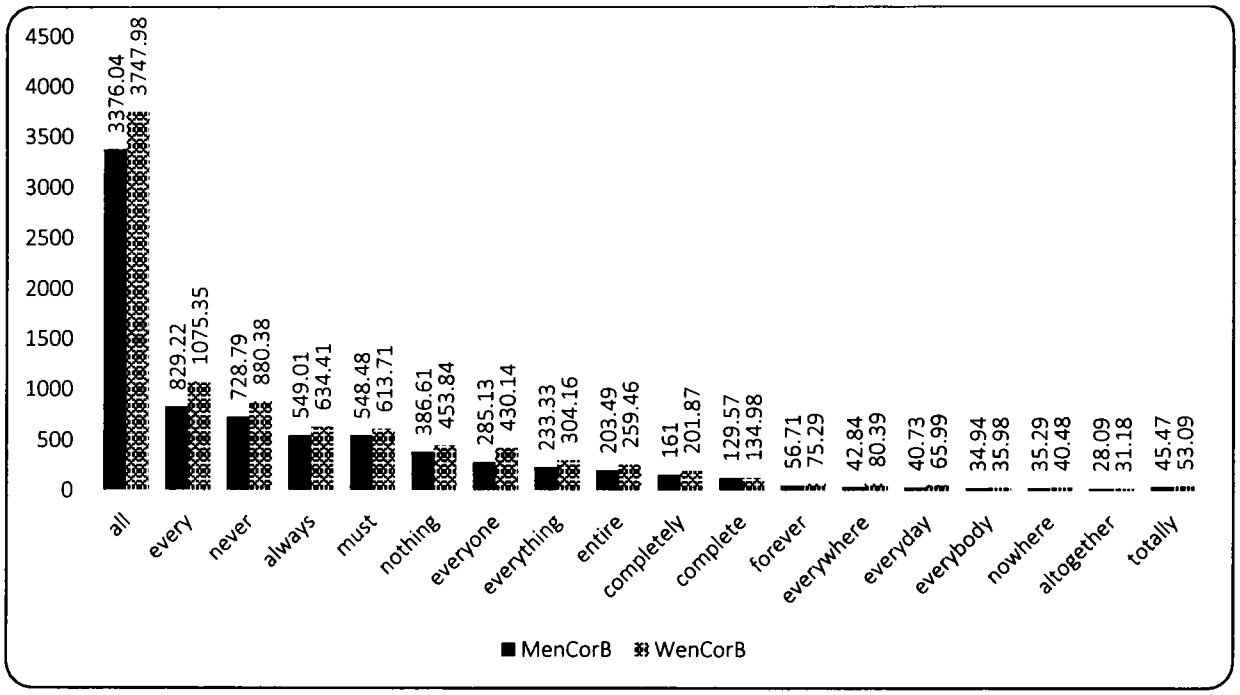


Figure 5.24. Certainty words used more by women (frequencies normalised to one million)

Differentiation

Despite extensive consultation of literature, no previous research could be found that focused exclusively on the use of words related to the linguistic category of *differentiation*. The results of the present study, therefore, are an important addition to language and gender studies. This category is conceptually similar in LIWC versions of 2007 and 2015 (Pennebaker, Boyd, Jordan, & Blackburn, 2015). Although research has been conducted on psychological processes as a whole, this variable has not been discussed exclusively. This category of LIWC2015 includes 81 expressions in all, e.g. *hasn't*, *but*, and *else*.

The Mann-Whitney U-test conducted for this study showed significant difference between genders on this variable. The results indicated that women used more expressions related to differentiation in comparison to men (U = 921412.500; U (standardized) = -4.211; Expected value = 1014444.000; Variance (U) = 488108265.841; p-value (Two-tailed) = < 0.0001).

5.6.4 Perceptual Processes

The broader category of words related to *perceptual processes* includes the total number of expressions that relate to the sub-categories of *seeing*, *hearing* and *feeling* (touch). Some of the previous research has found differences in men and women in the use of *perceptual processes*. For example, Yale (2007) studied college students' instant messaging (IM) to investigate gender differences in conversation dyads. The study found that girls used more words related to *perceptual processes* in comparison to boys. Similarly, Newman, Groom, Handelman, and Pennebaker (2008) examined a large sample of 14,000 texts and found similar results. However, using LIWC for their analysis, Brönnimann, Herlihy, Müller, and Ehlert (2013) studied transcribed testimonies of 24 victim witnesses and civil parties, which were translated from Khmer into English but they did not

find any significant difference between genders on the use of words related to perceptual processes. Similarly, Lenard (2016) analysed Congressional speeches but found no significant difference in the use of perceptual words between the genders.

The result of Mann-Whitney U-test conducted for this study were consistent with Yale (2007), and Newman, Groom, Handelman, and Pennebaker (2008). The analysis found significant differences between genders on the use of expressions related to this LIWC variable. The results indicated a tendency in women to overuse these words in comparison to men (U = 707021.500; U (standardized) = -13.915; Expected value = 1014444.000; Variance (U) = 488107958.197; p-value (Two-tailed) = < 0.0001).

Sense of Sight

This category of LIWC2015 is related to the language variable that covers words pertaining to humans ability to see objects around them. It includes expressions that relate to the names of colours, the things seen and the manner in which one can sees things. Some previous research (Kirk, 1992; Cameron, 2007, Newman, Groom, Handelman, & Pennebaker, 2008; and, Ardila, Rosselli, Matute, & Inozemtseva, 2011) did not report any statistically significant gender difference in the use of words related to this language variable. However, a study (Lenard, 2016) has reported that women significantly used more words related to the sense of sight in comparison to men.

The Mann-Whitney U-test conducted for the present study also found significant difference between genders. The results indicated that women tend to use expressions related to the sense of sight with far greater frequency than men (U = 820155.500; U (standardized) = -8.794; Expected value = 1014444.000; Variance (U) = 488086150.133; p-value (Two-tailed) = < 0.0001).

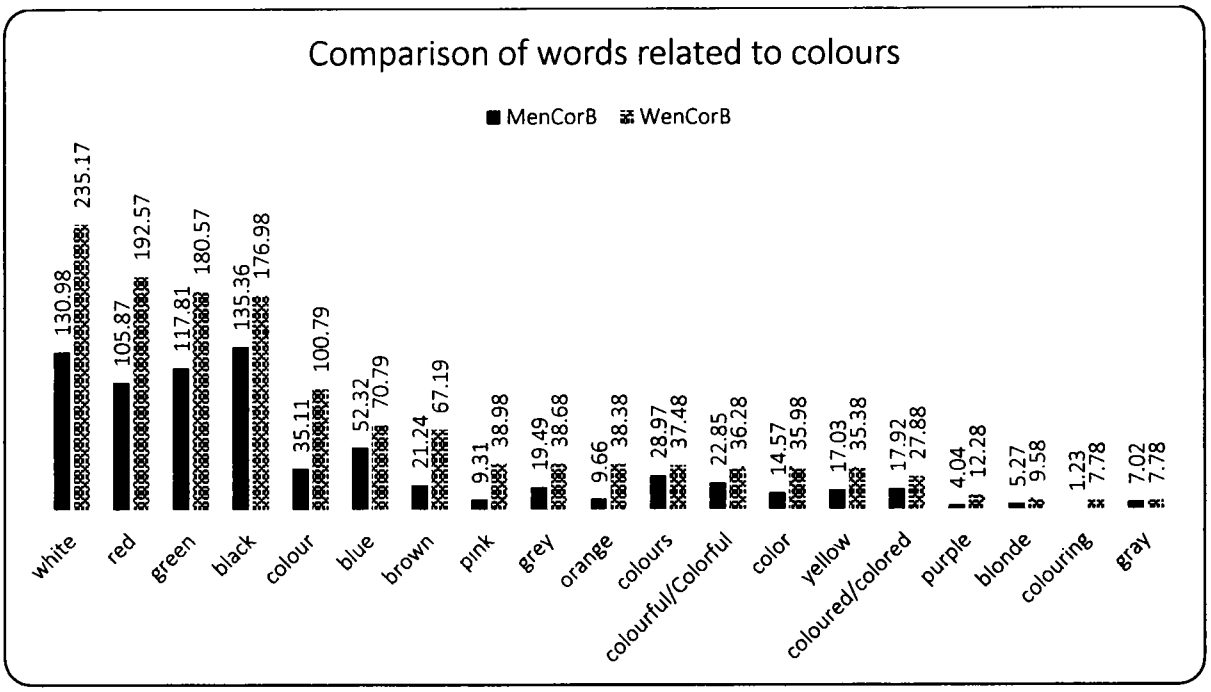


Figure 5.25. Words related to colours (frequencies normalized to one million)

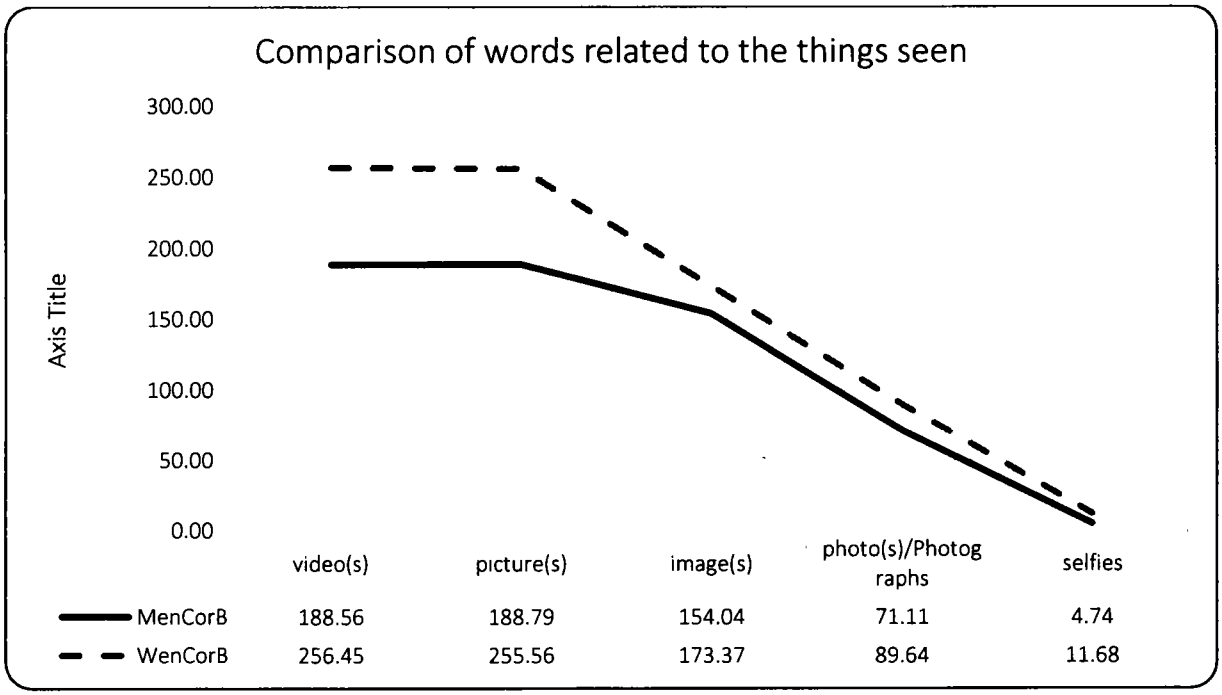


Figure 5.26. Words related to the things seen (frequencies normalised to one million)

After this analysis, the data was further explored to see gender differences in the use of words related to colours and the things seen. As shown in Figures 5.25 and 5.26, women were also found to use more words related to colours and the things seen (e.g. still images and movies).

Sense of Hearing

This category of LIWC2015 is related to the language variable that covers words, which relate to the sense of hearing such as the names of things heard (music, sound), the devices that produce sound (radio, telephone etc) and the manner in which sound is produced and heard (yell, hush, scream etc). Previous research by Yale (2007), and Newman, Groom, Handelman, and Pennebaker (2008) reported that females use words related to the sense of hearing more frequently than males. However, a study by Lenard (2016) did not find any significant difference between the two genders for this language variable.

The Mann-Whitney U-test for this study also found significant gender difference on this language dimension of LIWC2015. The results indicated that women used more words related to this language variable in comparison to men (U = 833549.500; U (standardized) = -8.189; Expected value = 1014444.000; Variance (U) = 487915185.802; p-value (Two-tailed) = < 0.00010.

Sense of Touch

Touch is the first sense humans develop, which is used to localize and recognize objects and get feedback from the environment. Furthermore, touch implies an important form of social presence. It plays a prominent role in human communication and is essential for the development of infants and the wellbeing of humans. Despite psychological significance of touch as a means of communication, not a great deal is known about when and why touching occurs and what meaning

it may have in the same-gender and opposite-gender interactions. Studies on gender and touch have focused on observed occurrences of touch, self-reports of touch frequency, people's beliefs about the meanings of touch, observers' perceptions of interactions that involved touch, and empirically assessed response to touch.

A limited research in the past, however, has focused on analysis of gender differences in the use of words related to the sense of *touch*. Studies by Kirk (1992), Yale (2007), and Newman, Groom, Handelman, and Pennebaker (2008) yielded consistent results and found that women used vocabulary related to the sense of touch more frequently than men. The Mann-Whitney U-test conducted for the present study also yielded results, which indicated that women used tactile-related words with far more frequency than men (U = 783448; U (standardized) = -10.458; Expected value = 1014444.000; Variance (U) = 487860013.254; p-value (Two-tailed) = < 0.0001).

5.6.5 Biological Processes

This category of LIWC2015 comprises words that relate to physical structure of human body, diet and health etc. Research on this category is extremely scarce in linguistics (Lenard, 2016). Therefore, the results of the present study might be valuable for future researchers should they decide to examine gender differences in the use of expressions related to biological processes category and its relevant subcategories. The only studies that could be found in the literature searched for this language category were conducted by Nagarajan and Hearst (2009), and Lenard (2016). The former study reported that men used more words related to the broader category of biological processes whereas the latter did not find any statistically significant difference in use of vocabulary related to biological processes. The Mann-Whitney U-test conducted for the present

study contradicted the results found by the previous research. The results for the present study showed that women tended to overuse words related to biological processes far more frequently than men (U = 628453.500; U (standardized) = -17.471; Expected value = 1014444.000; Variance (U) = 488108222.832; p-value (Two-tailed) = < 0.0001).

Body

This is the first sub-category of *biological processes*, which is composed of words related to body parts. The only previous study (Lenard, 2016) related to this sub-category, which was accessed for discussion in this section, found no statistically significant differences between the gender in the use of words related to body parts.

The Mann-Whitney U-test conducted for the present study, however, contradicted Lenard's (2016) results. This study found significant gender differences in the use of words related to this LIWC2015 variable. The results indicated that women used expressions related to body parts with far more frequency than men (U = 763547; U (standardized) = -11.358; Expected value = 1014444.000; Variance (U) = 487958160.572; p-value (Two-tailed) = < 0.0001).

After conducting hypothesis test for broader gender differences for this variable, the data was also quantitatively examined for subtle differences in the use of particular expressions that could not surface through Mann-Whitney U-test. For this purpose, words that occurred with a raw frequency of 10 or more for either gender were included in respective lists given in Table 5.7 and 5.8 below.

Table 5.7

List of words related to body parts used more by men (frequencies normalised to one million)

S No	Word	MenCorB	WenCorB	S No	Word	MenCorB	WenCorB
1	arm(s)	122.05	101.08	**13**	sensation	9.48	6.28
2	leg(s)	81.99	67.16	**14**	throats	7.02	5.08
3	beard	26.51	24.28	**15**	lung	9.48	3.90
4	bloody	25.63	23.68	**16**	muscular	5.79	3.59
5	injury	40.21	17.68	**17**	chests	3.34	3.00
6	injuries	23.18	15.88	**18**	joints	4.21	2.99
7	knee	13.34	9.88	**19**	nudity	4.04	2.40
8	muscle(s)	21.95	18.26	**20**	scalp	4.21	2.10
9	toes	7.90	7.78	**21**	bellies	2.46	1.50
10	veins	7.90	7.78	**22**	shirtless	1.93	1.50
11	bald	7.55	7.18	**23**	saliva	2.46	0.90
12	wrist	7.90	7.18	**24**	inhale	4.04	0.30

Table 5.8

List of words related to body parts used more by women (frequencies normalised to one million)

S No	Word	MenCorB	WenCorB	S No	Word	MenCorB	WenCorB
1	face	333.76	478.13	67	knees	7.20	12.88
2	eyes	189.97	344.65	68	palm	8.25	12.88
3	heart	200.50	308.06	69	cheeks	7.37	12.28
4	head	248.96	290.36	70	eyebrows	5.09	12.28
5	hand(s)	535.66	587.62	71	cheek	6.14	11.68
6	body	168.02	239.37	72	liver	3.86	11.08
7	blood	157.84	232.17	73	lungs	7.90	10.78
8	wear	52.32	165.88	74	vein	4.04	10.78
9	clothes	48.63	152.08	75	breastfeeding	5.09	9.88
10	eye	106.40	143.38	76	jaw	7.37	9.58
11	wearing	62.85	134.68	77	slept	6.50	9.58
12	skin	38.80	134.68	78	womb	1.58	8.98
13	faces	88.66	105.29	79	kidney	3.86	8.68
14	bodies	58.11	101.39	80	mouths	6.67	8.68
15	sleep	46.53	91.19	81	thirst	8.08	8.68
16	hearts	84.10	88.79	82	belly	5.79	8.38
17	feet	67.42	88.19	83	bleed	5.44	8.38
18	heads	66.19	76.79	84	thirsty	7.02	8.38
19	brain	40.73	69.59	85	palms	5.62	8.08
20	shoes	31.78	68.69	86	gut	5.62	7.78

Table 5.8 (Continued)

List of words related to body parts used more by women (frequencies normalised to one million)

S No	Word	MenCorB	WenCorB	S No	Word	MenCorB	WenCorB
21	mouth	35.11	62.99	87	hip	7.55	7.78
22	injured	44.42	61.79	88	pulse	1.93	7.78
23	fat	35.29	56.99	89	tooth	6.32	7.78
24	finger(s)	60.22	89.07	90	rash	3.69	7.20
25	breath	41.79	47.38	91	waist	5.09	6.58
26	shirt	22.82	45.58	92	elbow	1.93	6.00
27	foot	39.33	44.08	93	tummy	1.76	6.00
28	ears	26.69	39.28	94	spit	4.92	5.98
29	wore	16.15	37.78	95	noses	4.74	5.68
30	shoulders	32.30	37.18	96	bleeds	4.56	5.40
31	nose	16.68	35.98	97	nude	3.51	5.38
32	sleeping	18.61	35.68	98	thumbs	4.56	5.38
33	neck	21.95	32.08	99	genital	0.53	5.08
34	lips	15.63	31.48	100	thigh	1.58	5.08
35	stomach	17.56	31.48	101	bodily	1.93	4.80
36	throat	14.40	29.08	102	bodily	1.93	4.78
37	shirts	18.43	27.88	103	necklace	0.53	4.78
38	shoulder	22.47	26.38	104	spat	5.79	4.78
39	eyed	19.49	26.08	105	toothless	4.92	4.50
40	naked	18.61	26.08	106	eyeing	4.74	4.50
41	teeth	17.38	26.08	107	sleepy	4.21	4.50
42	chest	20.89	25.48	108	breasts	2.28	4.49
43	breast	12.11	24.88	109	eyebrow	1.58	4.48
44	wears	10.01	23.38	110	eyeing	4.74	4.48
45	breathe	14.92	22.48	111	sleepy	4.21	4.48
46	breathing	17.21	22.48	112	toothless	4.92	4.48
47	flesh	11.24	20.98	113	tongues	3.16	4.20
48	bleeding	15.80	20.68	114	elbows	1.23	4.19
49	brains	11.41	20.68	115	haircut	1.93	4.19
50	tongue	17.91	20.68	116	heel	2.98	4.19
51	shoe	13.17	20.38	117	limp	1.58	4.19
52	bone	12.99	18.28	118	necks	3.51	4.19
53	ear	15.80	17.98	119	thighs	1.05	4.18
54	facial	9.13	17.98	120	wrists	2.81	4.18
55	sweat	12.99	17.68	121	eyesight	1.93	3.89

Table 5.8 (Continued)

List of words related to body parts used more by women (frequencies normalised to one million)

S No	Word	MenCorB	WenCorB	S No	Word	MenCorB	WenCorB
56	nerve	10.53	17.38	122	genitals	0.53	3.89
57	spine	9.66	17.38	123	jaws	4.39	3.89
58	heels	5.09	16.18	124	urine	1.23	3.29
59	bones	11.76	15.88	125	uterus	0.18	3.29
60	lip	7.55	15.88	126	arousal	0.00	2.99
61	nerves	13.34	15.88	127	foreheads	1.93	2.99
62	asleep	8.25	13.78	128	nostrils	1.93	2.99
63	forehead	8.60	13.78	129	sweater	1.05	2.99
64	lipstick	1.58	13.48	130	hairy	2.11	2.70
65	guts	9.66	13.18	131	eyeballs	1.76	2.40
66	toe	9.66	12.90				

A general look over the 20 most frequent words in the two lists shows that both the gender used some gender-specific words: men used more words that related exclusively to masculine gender such as *beard*, *muscles*, *muscular, bald,* and *scalp* whereas women used more words that related to femininity, beauty and fashion, e.g. *face*, *skin*, *clothes*, *wearing,* and *shoes*.

Health

The issue of gender differences related to health has extensively been studied in social sciences and medical. Some of these studies (e.g. Umberson, 1992) have attempted to see the impact of social integration on men's and women's health and mortality with the conclusion that loneliness caused by widowhood had far greater adverse impact on the physical strength of men in comparison to women, ultimately leading to their death. Others (e.g. Ek, 2013) reported greater life expectancy in Finish women in comparison to men. Yet some others (e.g. Addis & Mahalik, 2003; Banks, 2001; O'Brien, Hunt, & Hart, 2005) have reported a typical tendency in men to avoid seeking health care as a mark of their masculinity. O'Brien, Hunt, and Hart (2005), however, added

that men sought health-care when such help was linked to performance of a more valued masculine task such as sex function or protecting others.

Research on language expressions used by men and women for health-related issues have also been studied in diversified contexts. For instance, a study in CMC context by Seale (2006) analysed postings by well-wishers of patients on breast- and prostate-cancer Internet forums. The research reported that women posted more frequently, even on prostate cancer forum, using an emotional communication style that is adopted by women generally. On the other hand, men posted with lesser frequency on breast cancer forums and experienced that an emotional style of expression did not suit their masculinity. Similarly, a corpus-based author profiling study (Soler, 2013) confirmed that words related to health were used more by women. In LIWC-based CL studies, however, Lenard (2016) did not find any statistically significant gender differences in the expressions related to the domain of health.

The Mann-Whitney U-test conducted for the present study also produced results that were in line with the past research. The results indicated that women used more words related to health more than men, which showed their concern for health (U = 750970.500; U (standardized) = -11.927; Expected value = 1014444.000; Variance (U) = 487968517.365; p-value (Two-tailed) = < 0.0001).

After conducting hypothesis test for this variable, the data was further analysed for gender differences in particular words used by men and women. Since a complete analysis of all the words captured by LIWC2015 for this variable was not possible, the words related to diseases and negative health habits (e.g. smoking, drinking alcohol) were studied. It was found that women mentioned names of diseases more than men except *diabetes*, which was used more by men. It was also found that men talked more about negative health habits and used the expressions *alcohol*, *cigar*, and *cigarettes* more in their blog posts in comparison to women.

The trends surfaced through this analysis in both the genders can also be corroborated by the findings of the previous research conducted in perspectives other than linguistics. For example, according to World Health Organisation (2016), in Pakistan men (10.0%) suffer from and have more risk factor of diabetes as compared to women (9.7%). This could be one of the reasons that men talk more about a disease to which they are more prone. Similarly, some previous research (e.g. Nathanson, 1977; Harrison, 1978; Waldron, 1988) claims that gender differences in health behaviour may arise from gender role socialization, which directs females' concerns toward health and safety and males' concerns toward competition, aggression, and risk-taking. Traditional gender role socialization not only encourages females to guard their own health and safety, but to be nurturing and attentive to the needs of others. If women assume these nurturing roles within marriage, they are likely to monitor their spouses' health and health behaviours as well as taking some responsibility for their spouses' health. There is some evidence that women are more likely than men to assume responsibility for the health of their spouses. Women are more likely to organize living habits – such as preparing food and monitoring health supplies and prescriptions – that can have an impact on the health of household members (Troll, Miller, & Atchley, 1979; Depner & Verbrugge, 1980). Marketing researchers are well aware of women's efforts to produce family health; when selling preventive health care, margarine, bran cereal, and other allegedly healthful products, advertisers pitch their ads toward women, who they assume make most health decisions for the family.

Although gender roles may steer men and women towards different health behaviour orientations, gender differences in negative health behaviour are not uniformly consistent across different kinds of health behaviours or cultures, particularly within certain age groups (Waldron, 1988). For example, in some cultures, women are as likely as men to smoke or drink.

Sexual

Previous research on gender differences in language use has paid little attention to the differences in men and women in the use of words related to sexuality. Two studies recently conducted in this regard have yielded contradictory results. One, conducted by Graells-Garrido, Lalmas, and Menczer (2015), studied gender bias in Wikipedia contents in the perspective of how members of the two genders are characterised in their biographies. For this purpose, the researchers selected the portion of Wikipedia that covered biographical content and conducted a three-dimensional analysis of the same: metadata, language, and network structure. One of their findings included the tendency in women to overuse words that related to sexuality. The second study was Lenard's (2016), which analysed political speeches by US congressmen and congresswomen. This study, however, found no significant differences between the genders in the use of words related to sexuality.

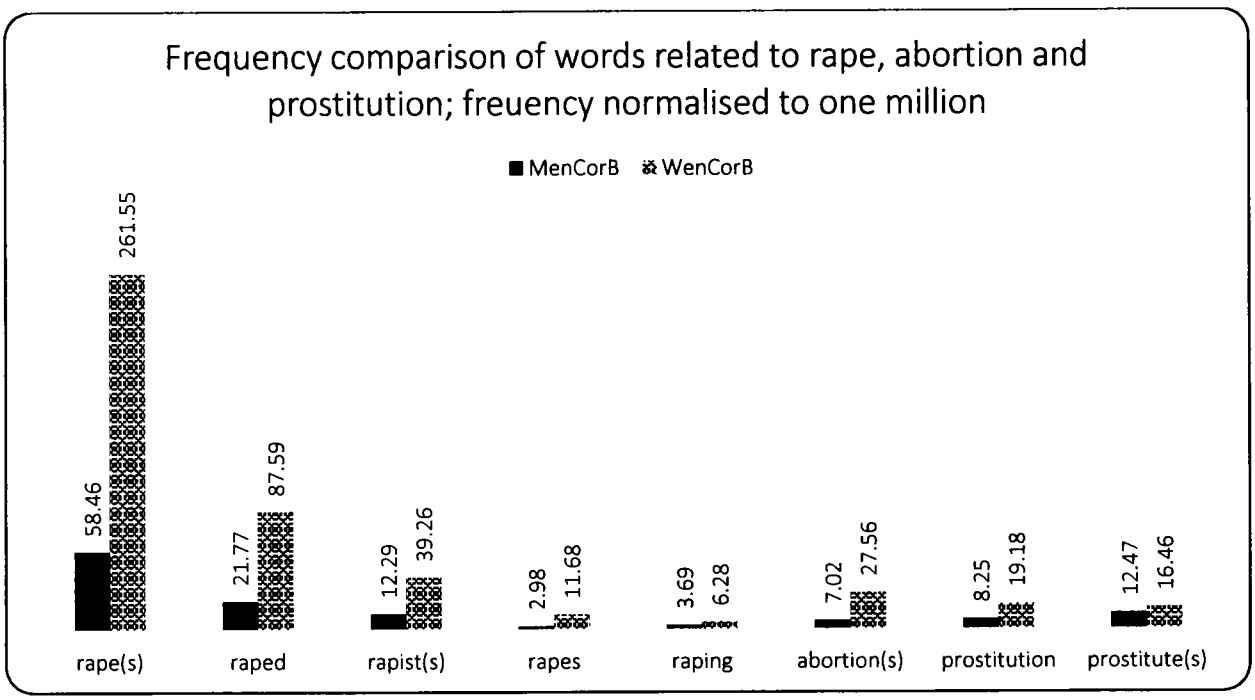


Figure 5.27. Words related to rape, abortion and prostitution (frequencies normalised to one million)

The Mann-Whitney U-test conducted for the present study also found significant differences between men and women in the use of words related to sexuality. The results were consistent with those of Graells-Garrido, Lalmas, and Menczer (2015) as women tended to overuse expressions related to this LIWC2015 variable (U = 894251.500; U (standardized) = -6.046; Expected value = 1014444.000; Variance (U) = 395158843.516; p-value (Two-tailed) = < 0.0001).

Since a more detailed analysis of all the words used by men and women for sexuality was not possible, words related to common social problems, namely, rape, abortion and prostitution were selected to find out gender differences in the use of their frequencies. As Figure 5.27 shows women used the words related to these aspects more in comparison to men. The reason for this is quite obvious: the words related to the problems faced by women and women bloggers could easily identify themselves with members of their own gender who faced these social problems.

Ingestion

In this category of LIWC2015, words are related to things that can be drunk or eaten, i.e. food items. Previously, limited research in linguistics has focused attention on studying gender differences in the use of terms related to *ingestion*. Within the available literature, only two studies were found, which had examined gender differences related to this aspect of language: One was conducted by Abbar, Mejova, and Weber (2015) who analysed the potential of Twitter to provide insight into US-wide dietary choices by linking the tweeted dining experiences of users to their interests, demographics, and social networks. They found that women tweeted about food, especially low-fat, more than men. The other study was of Lenard (2016), which analysed political speeches of US congressmen and congresswoman. This study, however, did not report any significant difference between the genders in the use of words related to ingestion.

The Mann-Whitney U-test conducted for the present study found significant differences in the use of expressions related to ingestion. The results indicated that women used more words overall in comparison to men (U = 821153; U (standardized) = -8.805; Expected value = 1014444.000; Variance (U) = 481921969.916; p-value (Two-tailed) = < 0.0001).

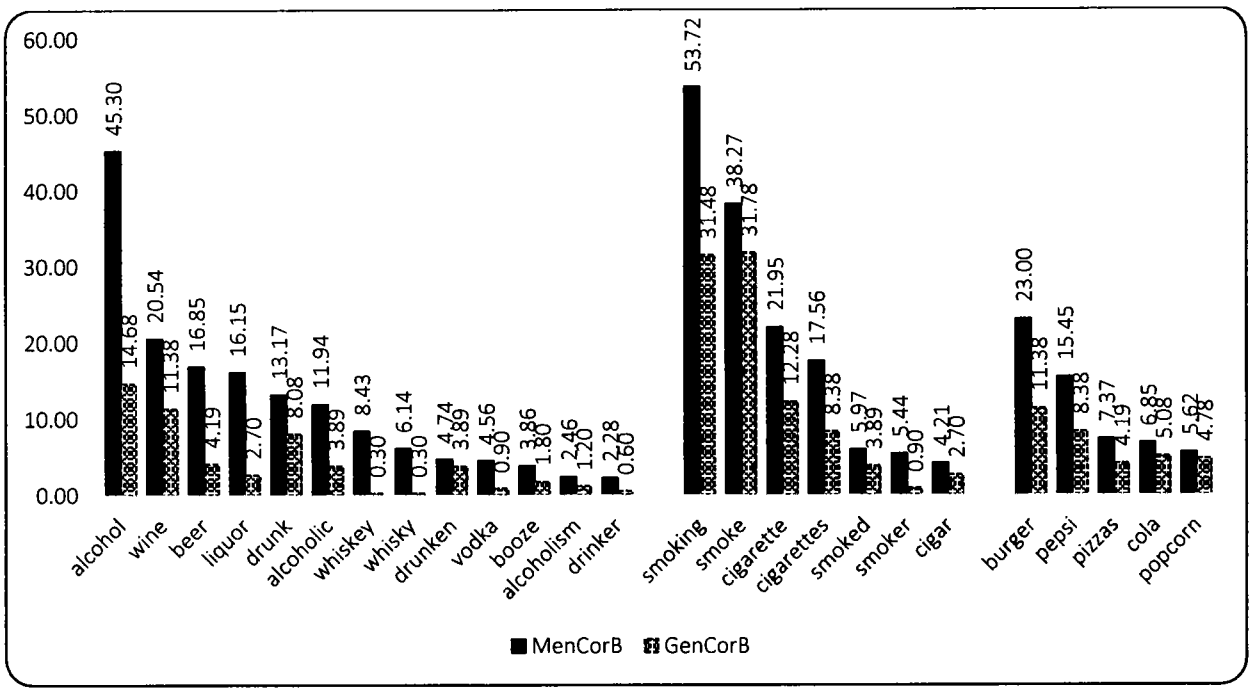


Figure 5.28. Words related to 'ingestion' used more by men (frequencies normalised to one million)

The data was further analysed for finding gender differences in particular words related to ingestion. While majority of the words in this category were used more by women, the words given in Figure 5.28 were used more by men. These words were related to the broader categories of alcohol, smoking, soft drinks and some of fast food items.

The overall findings of the present study are consistent with those of Abbar, Mejova, and Weber (2015) in linguistics. Also, gender differences in use of words related to the names of specific food items are in line with the studies conducted on gender differences in food consumption in the fields other than linguistics. For instance, a research in public health (Prattala, et al., 2007) found consistency in gender differences pertaining to food consumption patterns

across age, geography and educational level. The study found that gender-specific differences relate to both nutrient intake and recommendations for dietary allowances. Men have a higher energy intake and a higher percentage of the energy in men's diets is derived from animal products. Men's diets are characterized by meat, bread, and alcohol.

Previous research indicates that women have a higher awareness and better knowledge of nutrition than men. For women, nutrition frequently plays a central role in their conception of health. Women are more concerned about healthy eating habits (Kiefer, Rathmanner, & Kunze, 2005). Men prefer a traditional diet that is high in fat and meat, whereas women more frequently prefer healthy foods and a low-calorie diet (Missagia, Oliveira, & Rezende, 2013). Similarly, a study on children between 11 to 14 years age also showed that male preferred pizza, soft drinks and packed snacks in their daily diet more than female at different times of a day (Cammarota, Laurino, & Pellicano, 2008). Studies in health also indicate that men smoke more than women on daily basis (Allen, Scheurmann, Nollen, Hatsukami, & Ahluwalia, 2016).

5.6.6 Drives

This LIWC2015 category includes words expressing human drives related to its sub-categories of *affiliation*, *achievement*, *power*, *reward* and *risk*. Being a new category in LIWC2015 version, this general category has not received much attention by linguists so far. Therefore, the results of the present study might be valuable for future researchers should they decide to examine gender differences in the use of expressions related to human drives as a whole. The Mann-Whitney U-test conducted for the present study found significant gender differences in the use of expressions related to this category. The results for the present study showed that men tended to overuse words related to overall human drives far more frequently than women (U =

1124428; U (standardized) = 4.978; Expected value = 1014444.000; Variance (U) = 488114795.552; p-value (Two-tailed) = < 0.0001).

Affiliation

The first sub-category within the broader category of 'drives' is *affiliation*. Gender differences have often been characterized by at least one of two dimensions: (1) affiliation and interpersonal warmth versus impersonality and coldness, and (2) assertiveness and dominance versus indirectness and passivity. These two dimensions, *affiliation* and *assertiveness*, are common in language studies. The study of gender differences in expressions of affiliation drive has been investigated by some research both in the fields of linguistics and psychology. In linguistics, Leaper and Ayres (2007) summarized decades of research by organizing meta-analyses of gender-linked language around the interpersonal dimensions of affiliation and assertiveness. They defined assertive language as the language used to influence, such as imperative statements, suggestions, criticisms, and disagreements. Affiliative language was defined as language affirming the speaker's relationship with the listener, including statements of support, active understanding, agreement and acknowledgment. The meta-analysis indicated that men used more assertive language and women used more affiliative language, but the sizes of these differences were moderated by methodological features of each study. For example, differences in assertiveness were most pronounced when participants were asked to discuss non-personal topics or to deliberate a specific issue. DeYoung, Weisberg, Quilty, and Peterson (2013) described the trait of affiliation as a tendency towards "enjoying and valuing close interpersonal bonds and being warm and affectionate" (p. 314).

The Mann-Whitney U-test conducted for the present study showed significant gender differences in the use of expressions related to affiliation. The results indicated that women tend to use words related to the variable of affiliation more than men (U = 873111.500; U (standardized) = -6.397; Expected value = 1014444.000; Variance (U) = 488111606.785; p-value (Two-tailed) = < 0.0001).

Table 5.9

Specific nouns used by women for expressing 'affiliation' (frequencies normalised to one million)

Dimension	MenCorB	WenCorB
Family		
famil(ies)	534.79	1080.15
parent(s)	196.64	539.92
husband(s)	67.59	381.54
wife/wives	162.40	277.76
brother(s)	205.94	224.37
relationship(s)	112.36	208.17
sister(s)	89.91	206.97
wedding(s)	56.01	194.97
relatives	38.10	81.59
cousin	26.16	65.39
partner	33.01	46.18
siblings	17.73	39.28
Friends		
friend(s)	453.32	850.08
colleague(s)	54.60	83.07
gang	31.60	44.38
Love Relation		
love	402.06	692.60
loving	29.32	49.79
valentine	21.95	31.48
lovers	25.63	30.28

Table 5.9 (Cont'd)

Specific nouns used by women for expressing 'affiliation' (frequencies normalised to one million)

Dimension	MenCorB	WenCorB
Socialisation		
dance	47.05	73.79
dancing	33.53	48.58
celebration	22.82	32.98
Communication		
message(s)	146.60	216.57
conversation(s)	80.06	141.27
communication	37.22	41.68
Social Media		
facebook	168.72	232.47
twitter	90.77	95.39
Affiliative Feelings		
help	406.62	527.93
sympathy	25.63	29.08
compassion	16.85	29.98
Common Identity		
we	4201.74	5289.16
our	2780.50	3200.25
us	1696.89	1902.94
ourselves	109.91	157.78
ours	36.69	44.38

Table 5.10

Specific nouns used by men for expressing 'affiliation' (frequencies normalised to one million)

Dimension/Word	MenCorB	WenCorB
Alliance		
alliance	49.34	17.68
association	45.12	34.18
allies	35.64	12.58
Brotherhood		
brotherhood	28.27	11.08

Table 5.10 (Continued)

Specific nouns used by men for expressing *'affiliation'* (frequencies normalised to one million)

Dimension/Word	MenCorB	WenCorB
Group(s)		
group(s)	515.12	394.15
league	270.55	52.49
club(s)	201.73	81.27
union	71.98	30.58
Team(s)		
team(s)	1177.72	284.05
patry/parties	1008.65	464.34
game(s)	628.54	195.56
member(s)	365.54	266.96
relations	103.59	55.49
partnership	25.81	10.48
Military Members		
squad	98.32	16.78
troops	38.10	20.10
Affiliative Feeling		
cooperation	27.74	10.48
Socialisation		
gathering	37.22	29.98

A further analysis was made to distinguish between the specific nouns used by both the genders for expressing *affiliation*. For this purpose, the top 100 most frequent words used in each of the corpora were analysed and lists of nouns were extracted for each corpus. Subsequently, the nouns were grouped under similar arbitrary categories. It was found that women used more affiliation expressions that related to family, friends, love relation, socialisation, communication, social media, affiliative feelings, and common identity. Men, on the other hand, were found to overuse affiliation expressions that related to alliances, brotherhood, grouping, teams, military bodies, socialisation, and affiliative feelings. Of special attention is the difference in the terms related to socialisation and affiliative feelings. In the former, women used words that related to dancing and celebrations whereas men used terms related to social gatherings. In the latter, women

used terms that expressed the feelings of sympathy, compassion and help, whereas men used vocabulary related to cooperation.

Achievement

Previous research on gender differences in the use of words related to achievement has produced mixed results. One of the earliest studies that examined gender differences in the distribution of words related to achievement words (Thorne & Henley, 1975) found that men preferred topics that related to achievement and work more in comparison to women. Later, analysing children's beliefs and responses to failure and success in mathematics, another study (Stipek & Galinsky, 1991) also reported that boys felt pride and achievement more in comparison to girls. However, this could also be attributed to boys' outperformance of girls in mathematics in general.

In contrast, another study (Schmader, Whitehead, & Wysocki, 2007) conducted a LIWC-based examination of recommendation letters with the hypothesis that those written for females would contain less achievement and more communication skills references. The study did not support the hypothesis since it did not find statistically significant differences on the use of achievement words. Some recent studies (Adler, 2013; Lenard, 2016) have quite different findings and argued that women were more likely to report pride of their achievements, thus indicating possible changes in linguistic choices.

The Mann-Whitney U-test conducted for the present study also found significant gender differences in the use of terms related to achievement. The results indicated that men used more achievement words with far more frequency than women (U = 1220849; U (standardized) = 9.342; Expected value = 1014444.000; Variance (U) = 488104259.656; p-value (Two-tailed) =

< 0.0001). The results of the present study were consistent with those of Thorne and Henley (1975), and Galinsky (1991).

Power

Mulac and Bradac (1995) argue that relationship between gender, language and power is much more complex than can be understood at any time, as power interfaces with both gender and discourse. Traditionally, women's language has been interpreted as 'powerless'. For instance, Tannen (1990b) argues that men's discourse has assertive and competitive features, whereas wome's is supportive and relational, leading to the distinction between male "report talk" and female "rapport talk". The same feature has been reported by Colley and Todo (2002) who examined email messages and found that e-mails from female participants contained a higher incidence of features associated with the maintenance of rapport and intimacy than those from male participants. Rosseti (1998) has also found that males are more prone to write in an aggressive, competitive style, while women tend to be far more supportive in their writing (email messages). It seems that male/female language style dichotomy has been transported into computer communication as well.

Trudgill (1977) has also speculated that women's overt prestige orientation was a result of their powerless position in society. He argued that in as much as society does not allow women to advance their power or status through action in the marketplace, they are thrown upon their symbolic resources, including their appearance and their language, to enhance their social position (Eckert, 1989). In the same line of research, Erickson, Lind, Johnson, and O 'Barr (1978) utilized the terms "powerful speech style" and "powerless speech style" and argued that speech style is linked to social power and status. Low-status persons generally use a powerless speech style laced

with intensifiers ("so", "very"), hedges ("I think", "kinda"), hesitations ("uh", "well"), hypercorrect grammar, questioning forms (use of rising question intonation in declarative form), polite forms, and gestures. High-status persons rarely use these powerless forms and, therefore, employ what Erickson, Lind, Johnson, and O 'Barr (1978) dubbed the powerful style.

The Mann-Whitney U-test conducted for the present study also found significant gender differences in the use of words related to power. The study found that men used these words far more frequently than women (U = 1259779.500; U (standardized) = 11.105; Expected value = 1014444.000; Variance (U) = 488113104.063; p-value (Two-tailed) = < 0.0001). Thus, the results are consistent with the previous research.

Reward and Risk

Reward and risk are two separate variables in LIWC2015 category but they have been treated together in this sub-section in line with some of the previous research. While gender differences pertaining to these two human drives have been extensively researched in the field of psychology and other social sciences, previous studies in the field of linguistics with focus on gender differences in the use of words related to reward and risk appear to be extremely rare. Many researchers in fields other than linguistics agree that women are more risk averse than men. For instance, after analysing 150 studies from 1967 to 1997, Byrnes, Miller, and Schafer (1999) concluded that female responders are more risk averse than their male counterparts. However, after analysing different studies using Cloninger's tri-dimensional personality questionnaire (TPQ) across a variety of Asian, European and Western cultures, Li, Huang, Lin, and Sun (2007) concluded that, taken together, gender differences in the personality traits depend somewhat on cultures, with women consistently demonstrating greater harm avoidance and reward dependence

than men in the Western world. Especially, they reported, that among Asian (Taiwanese) people, men and women did not differ in any of the Cloninger's personality dimensions.

If the role of culture is considered to be crucial for gender differences in reward and risk-taking drives, the results of the present study could be claimed to have provided insights from the field of linguistics towards the conclusions drawn by Li, Huang, Lin, and Sun (2007). The results of the present study are, therefore, consistent with those of the TPQ research conducted on Taiwanese people. The present study also did not find any significant difference in the use of words related to the linguistic categories of reward and risk given in LIWC2015. However, further research in linguistics will be needed before any conclusive statement could be made.

5.6.7 Time Orientation: Focus on Past, Present and Future

Under the broader LIWC2015 category of *time orientation* are included three sub-categories of *past focus*, *present focus* and *future focus*. The three sub-categories are grouped here for convenience in discussion.

Since the 1950s, the time perspective has led to numerous studies in the fields of psychology and education. Although the accumulated knowledge about the time perception is based on quite varied results, two closely linked time concepts can be differentiated (Holman & Silver, 1998): the *time perspective* and the *time orientation*. The former is characterized by the cognitive distance of the goals along with other properties such as their degree of coherence, number, affectivity or continuity. The latter refers to each person's differential tendency to be focused on the past, present or future, and where our conception of it lies.

The time orientation is conceived as a process lying at the origin of individual and social behaviour that is often unconscious and makes it possible to decompose and organize the

continuous flow of behaviour into different time frames, giving it meaning and coherence (Zimbardo & Boyd, 2008). This psychological time construction arises from the cognitive processes that allow differentiation between the past, present, and future, and has a strong influence on behaviour. The future time perspective has been highlighted as allowing people to set goals and regulate their behaviour to reach them (Suddendorf & Corballis, 1997; Roberts, 2002).

With regard to gender and time dimension, the results are scarce (Zalesky, 1994). In general, majority of studies on time orientation find no gender differences (Bouffard, Bastin, & Lapierre, 1994; Lapierre, Bouffard, & Bastin, 1997). However, some research (Zimbardo & Boyd, 1999) reported that females are more oriented towards the positive past and the future and others (Newman, Groom, Handelman, & Pennebaker, 2008) found no gender differences in overall time-orientation but reported that women used more words related to past and present time with no gender differences in words related to future time.

The Mann-Whitney U-test conducted for the present study found no statistically significant gender differences in the use of expressions related to past and future focus. However, gender differences were found for present focus. The hypothesis test conducted for this variable showed that, in comparison to men, women used more words that related to focus on the present (U = 807516; U (standardized) = -9.366; Expected value = 1014444.000; Variance (U) = 488114892.658; p-value (Two-tailed) = < 0.0001).

Zimbardo and Boyd (2008) differentiate among three types of present-orientation: present hedonism, present fatalism, and holistic present. While the former two are negative present-orientation, the last one is a balanced orientation towards present. Holistic present has also been termed as 'mindfulness', which can be defined as the degree of awareness that is achieved by purposefully paying attention to the present moment, without judging it (Kabat-Zinn, 1994). Since

the launch of the first empirical research conducted on this topic (Kabat-Zinn, 1982) numerous studies have shown the positive effects of increased mindfulness—on life-satisfaction, vitality, self-esteem, empathy, optimism, integrity or positive affect—and its contribution to reducing the difficulties with emotional dysregulation, depression, neuroticism, rumination, social anxiety, and wandering thoughts (Brown & Ryan, 2003; Thompson & Waltz, 2007; Dekeyser, Raes, Leijssen, Leysen, & Dewulf, 2008; Keng, Smoski, & Robins, 2011; Rasmussen & Pidgeon, 2011).

There is also very little research looking at whether males and females report similar or different levels of mindfulness. In most of the existing studies, gender differences in mindfulness are not found (Brown & Ryan, 2003; MacKillop & Anderson, 2007; Catak, 2012). However, some research has reported higher levels of mindfulness among females than males (Tamres, Helgeson, & Janicki, 2002; Bryant, 2003). The findings of the present study provide evidence from the perspective of linguistics for the psychological questions on gender differences in time orientation.

5.6.8 Relativity

This broader LIWC2015 category includes the sub-categories of *motion*, *space*, and *time*. Very limited research could be found in the available studies that focused on gender differences in words expressing *relativity*. For instance, studying the category in general in their profiles' study, Nagarajan and Hearst (2009) reported that women use words related to relativity far more frequently in comparison to men. In contrast, Lenard (2016) did not find any significant gender differences in the use of words related to this dimension.

The Mann-Whitney U-test conducted for the present study found significant gender differences in the use of expressions related to relativity. The results showed that men used

relativity words with far more frequency than women (U = 1095282; U (standardized) = 3.659; Expected value = 1014444.000; Variance (U) = 488115008.161; p-value (Two-tailed) = 0.0001).

Motion

Previous research in linguistics on the study of gender differences in words related to motion is extremely rare. In their research analysing 14,000 text samples, Newman, Groom, Handelman, and Pennebaker (2008) reported that women used motion verbs significantly more than men. Similarly, Isaac, Chertof, Lee, and Carnes (2011) who studied evaluation letters of 227 male and 70 female medical students found that female students used motion-related words with greater frequency in comparison to their male counterparts.

The Mann-Whitney U-test conducted for the present study found no significant gender differences in the use of words related to motion (U = 981590.500; U (standardized) = -1.487; Expected value = 1014444.000; Variance (U) = 488098858.294; p-value (Two-tailed) = 0.137).

Space

The controversy over differences between men and women in spatial ability has been existing in the field of psychology since Maccoby and Jacklin (1974) reported that female outperform men in verbal abilities whereas men outperform women in spatial abilities. Some studies (e.g. Sander, Soares, & D'Aquila, 1982) observed high effect size (.80) for such trend in men on special tasks like mental rotation of 3-D figures. While some research Casey, Nuttall, Pezaris, and Benbow (1995) has highlighted the practical pedagogical implication of such results in teaching mathematics, others have attributed the existence of this difference in genders to both biological (Kimura, 1992) and sociological (Regian & Shute, 1993) factors. Later, consistent

results were reported by other studies conducted on human cognition and psychology in different contexts such as mental rotation test (Delgado & Prieto, 1996; Parsons, Larson, Kratz, Thiebaux, Bluestein, Buckwalter, & Rizzo, 2004; Driscoll, Hamilton, Yeo, Brooks, & Sutherland, 2005; Iachini, Sergi, Ruggiero, & Gnisci, 2005); cognitive spatial tests for 90 men and 104 women on navigation strategies and knowledge about geographic locations on local and world maps (Dabs, Chang, Strong, & Milun, 1998); the role of right hemisphere on spatial tasks performed by 104 right-handed males and females (Rilea, Roskos-Ewoldsen, & Boles, 2004); and, cognitive performance by children of different ages (Ardila, Rosselli, Matute, and Inozemtseva, 2011). These results reinforced that men score higher on spatial ability measures in comparison to women.

Such research in psychology also encouraged some research in linguistics to study if gendered language choices also testify to similar trends in men and women. For example, LIWC-based studies on instant messages by college students (Yale, 2007) and political speeches by US congressmen and congresswomen (Lenard, 2016) reported that males scored higher on spatial ability in comparison to females. In the same vein, a study on 297 Medical Student Performance Evaluation (MSPE) letters for 227 male and 70 female medical students (Isaac, Chertof, Lee, and Carnes, 2011) also confirmed that men used more spatial references than women. While some of the research (Newman, Groom, Handelman, & Pennebaker, 2008) found greater mean rank among men for this category, it did not report such difference as significant. However, none of the studies accessed for discussion in this section reported that women showed more spatial ability than men.

The Mann-Whitney U-test conducted for the study found significant gender differences in the use of the expressions related to space. The results showed than men used these expressions far more frequently than women (U = 1147584.500; U (standardized) = 6.026; Expected value =

1014444.000; Variance (U) = 488113307.899; p-value (Two-tailed) = < 0.0001). The results of the present study were also in line with those of the previous research.

Time

Time factor has been considered in studies on deception and honesty to see if people are consistent in their statements or change their words over a period of time (e.g. Pennebaker, Mehl, and Niederhoffer, 2003). There is, however, a limited research available in language difference between men and women with regard to temporal-orientation. Among such studies, Iosub, Laniado, Castillo, Fuster Morell and Kaltenbrunner (2014) applied LIWC and SentiStrength as tools to examine a community of 12,000 Wikipedia editors, focusing on how their gender, status and network medium influenced their language choices. On the basis of difference attributed to gender, they reported that time-related words (past and present) were used by female editors with greater frequency than male editors. However, Newman, Groom, Handelman, and Pennebaker (2008) did not report any significant difference between genders on the category of time.

In contrast to the findings of these studies, the Mann-Whitney U-test conducted for the present study yielded consistent results to Newman, Groom, Handelman, and Pennebaker (2008). The results showed no statistically significant gender differences in the use of words related to time (U = 1016336.500; U (standardized) = 0.086; Expected value = 1014444.000; Variance (U) = 488112350.855; p-value (Two-tailed) = 0.932).

5.6.9 Personal Concerns

The general category of *personal concerns* is not calculated by LIWC2015. Rather this general category consists of six sub-categories of *work*, *leisure*, *home*, *money*, *religion*, and *death*.

The sub-categories also present different themes. An analysis of these sub-categories is presented in the succeeding relevant sub-sections.

Work

An important contribution to gender differences in topic choice was made by Moore (1922) who theorised that these differences were universal as they were biologically determined. However, after a long span of seven decades, Bishoping (1993) conducted a test of Moore's ideas and compared his findings to similar research carried out between 1922 and 1990. However, Bishoping's (1993) findings were not consistent with those of Moore's (1922). She found that, over the time, work-related topics have increased in women's language in comparison to men's. She concluded that topic-choice could thus be not attributed to biological determinism. Despite a relative reduction of gender differences in work-related topics, men still prevailed in their use, which was confirmed by Fehr (1996) who believed that, being a non-personal topic, work was the men's choice. In addition to these studies, some research (e.g. Lenard, 2016) has not reported any statistically significant gender differences in the use of the expressions related to work.

The Mann-Whitney U-test conducted for the present study found significant gender differences in the use of words related to work. The result showed that men used these terms more in comparison to women (U = 1264872.500; U (standardized) = 11.335; Expected value = 1014444.000; Variance (U) = 488113427.058; p-value (Two-tailed) = < 0.0001).

A further analysis was carried out to see any differences in the use of words by both the genders that could indicate for possible differences in theme choice. For this purpose, top 100 most frequent words from the expressions captured by LIWC2015 in both the corpora were compared

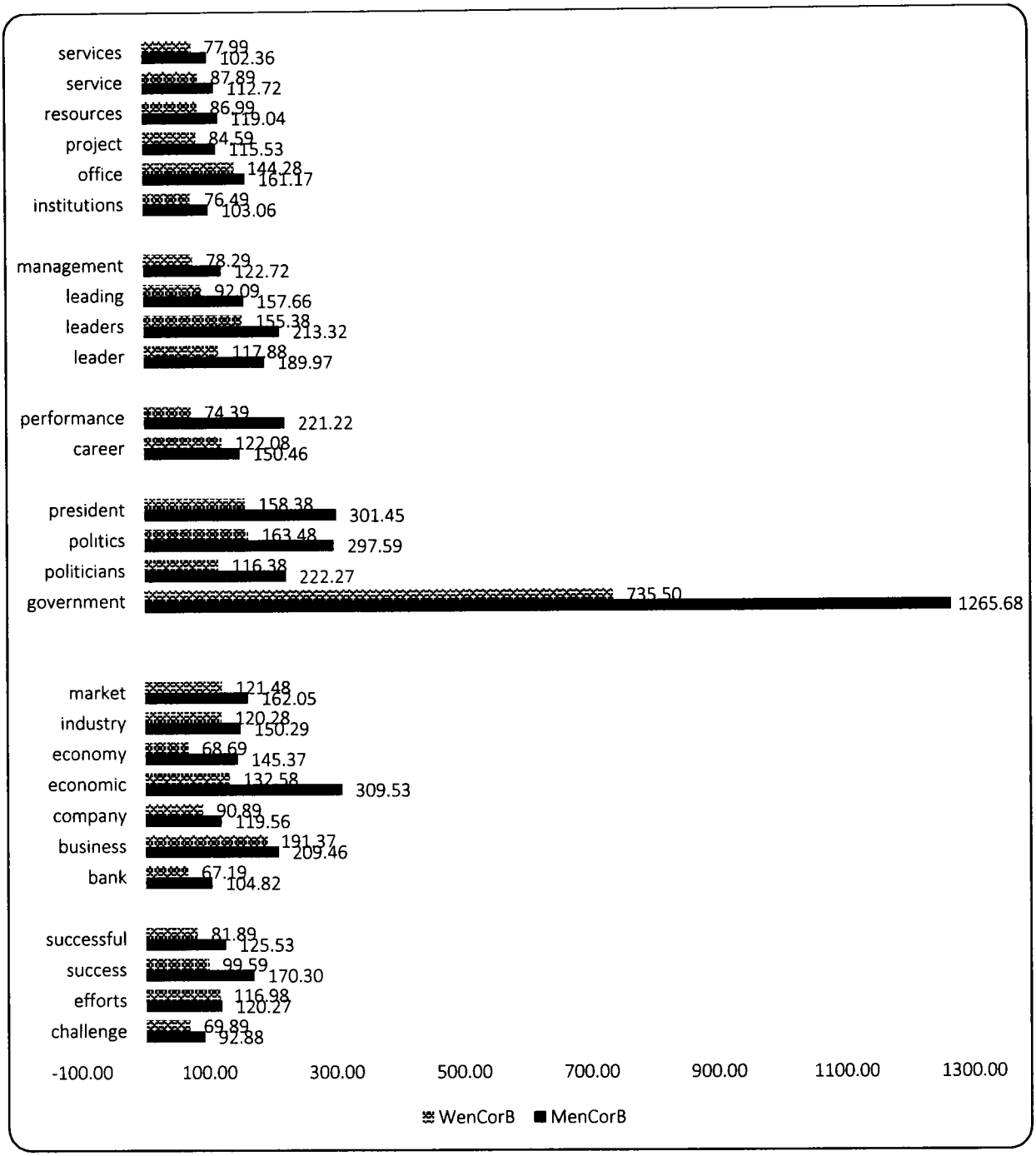


Figure 5.29. Words related to 'work' used more in MenCorB (frequencies normalised to one million)

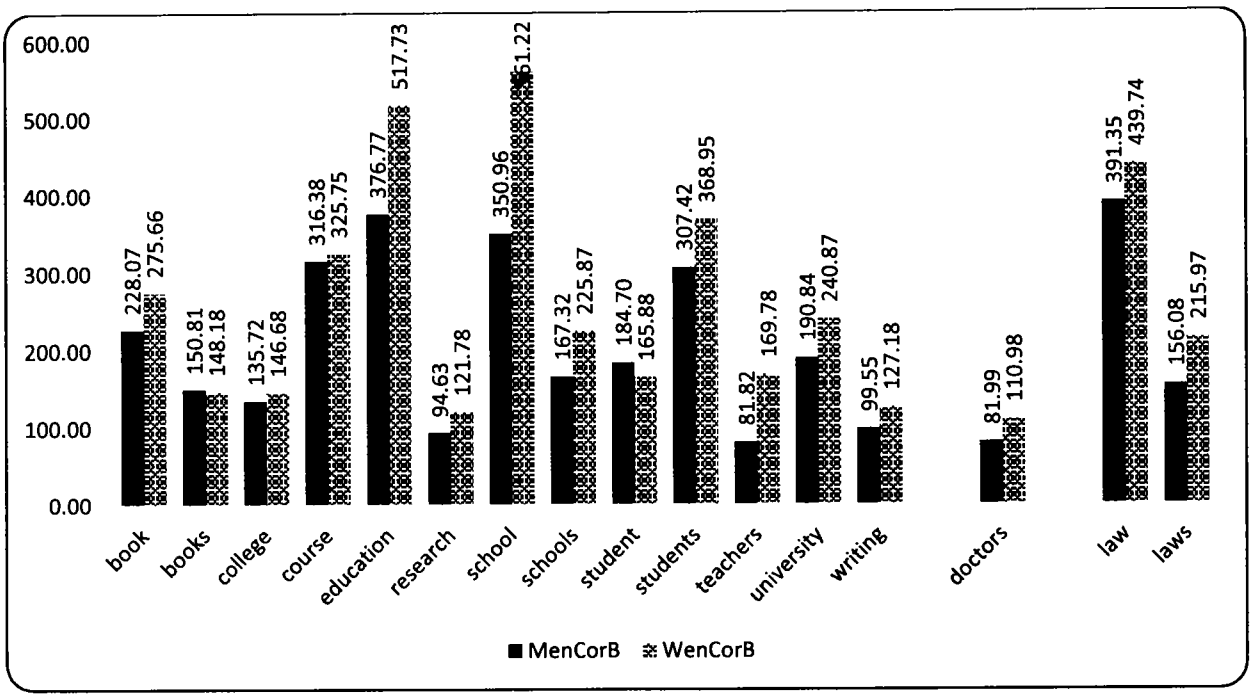


Figure 5.30. Words related to 'work' used more in WenCorB (frequencies normalised to one million)

for the obvious reason of difficulty in comparing all the words in both the corpora found by LIWC2015. The words were then grouped thematically as shown in Figures 5.29 and 5.30. It was found that, in work category, men used words that related to the themes of job, politics, success/failure, business/industry, and leadership/management. On the other hand, women tended to use more words related to the themes of education, health and law. A possible explanation for these differences may be found in the different social roles that men and women play in Pakistani society with men's increased participation in jobs, politics, business, industry and other challenging areas. The overuse of words related to education by women is a positive social indicator of their social uplift.

Leisure

An area of practical and social significance is the involvement of men and women in leisure activities. Previous research about gender differences has provided contradictory results. Since

Moore's (1922) study, this aspect of language research has been given systematic treatment in research. Later research also established gender differences in leisure preferences (e.g. Gentry & Doering, 1979; Gruber, 1980) and participation (Bishop & Witt, 1970; Ragheb, 1980). The meta-analysis of research from 1922 to 1990 conducted by Bischoping (1993) indicated that men preferred talking about leisure activities in general and sports in particular. However, the same study also observed gradual rise in leisure preferences by women. Similarly, Eggins and Slade (1997) observed conversation during coffee breaks at workplace and reported that men talked more about sports whereas women shared personal experiences. Later, Martin Rojo and Gomez Esteban's (2005) research yielded similar results. They, however, added that men did not feel at ease when talking about their personal experiences but were comfortable when they talked about sports. The results that men talk more about sports or leisure activities in general were further confirmed even by more recent studies such as Yale (2007), Newman, Groom, Handelman, and Pennebaker (2008), and Manjavacas (2015). In contrast, Lenard (2016) did not record any statistically significant differences in language about leisure activities in general.

The Mann-Whitney U-test conducted for the present study showed no statistically significant differences in the use of language about leisure activities in general (U = 1039739.500; U (standardized) = 1.145; Expected value = 1014444.000; Variance (U) = 488098482.056; p-value (Two-tailed) = 0.252). After conducting the hypothesis test for studying any differences in the overall use of words related to leisure activities, a detailed analysis was carried out to see any difference between men and women in themes pertaining to leisure activities. For this purpose, all the words captured by LIWC2015 with a raw frequency of 10 in either corpus were analysed. The words were later grouped under arbitrary themes related to leisure activities as shown in Table 5.11.

Table 5.11

Gender differences in themes related to 'leisure'

MenCorB	WenCorB
Alcohol/Drinks: alcohol, alcoholic, drinks, drunken, drinking, drunk, wine, beer	**Art:** Artist, artists, art
Movies/Film: blockbuster, DVD, YouTube, CDs, cinemas, cinematic, cinematography, comic, comical, filmmaker, filmmakers, film, films, theatre, theatres, movie, movies, cinema, comedy, DJ	**Movie/Films:** theatre, drama, videos
Music: concert, concerts, music, musician, musicians, guitar, sung, sang, danced	**Music:** Dance, dancer, dances, dancing, singing, sings, song, songs, sing, musical
Literary Readings: poetry, novels	**Food/Cooking**: Cafes, Café, cafeteria, cooked, coffee, cookie, cookies, cooking, cooks, cook, chilling
Entertainment: entertain, entertaining, actor, actors	**Entertainment:** Comedian, actress, actresses, amused, amusement, amusing, celebrities, celebrity, comics, dramas, entertained, entertainment, fun, joked, jokes, joking, joke
Indoor Games: Chess, cards	**Social Media:** Blog, chat, blogger, bloggers, blogging, blogs, chatted, chatter, chatting, selfies, facebook, Instagram, twitter, selfie,
Sports/ Outdoor Games: Bowling, coach, coaches, coaching, cricket, Djokovic, hockey, ball, football, game, games, golf, soccer, sport, sporting, sports, tennis, team, swim	**Socialisation:** Partying, celebrate, celebrated, celebrates, celebrating, celebration, celebrations, celebratory, festive, festivities, festivity, families, family, festival
Physical Fitness: Running, yoga, nap, play, fitness, played, playing, plays	**Rest/Relaxation:** Holiday, Holidays, weekend, vacation, vacations, weekends, relaxation, relaxed, relaxing, dreaming, fantasies, dreamed, dreams, dreamy, relax, dream
Travelling: **Travels, museum, hotel, booked**	**Travelling:** Travel, beach, beaches, hotels, restaurant, travelled, traveling, restaurants
	Shopping: Shop, shoppers, shopping, shops

It was observed that both men and women used words pertaining to themes like *travelling*, *music* and *entertainment*. However, men used words that related to themes like *alcohol, movies/films*, *literary readings*, *indoor games, sports/outdoor games* and *physical fitness*. Women,

on the other hand, used words that could be grouped under the themes of *art, food and cooking, social media, socialisation, rest/relaxation* and *shopping*.

Home

The interest in gender differences in the use of vocabulary related to home has recently been shown by a few studies. All these studies were LIWC-based, which yielded consistent results. For instance, analysing a corpus of 14,000 texts from a variety of genres, Newman, Groom, Handelman, and Pennebaker (2008) reported that women used more words related to home in comparison to men. Similarly, in non-English domain of research in sociolinguistics, Manjavacas (2015) analysed Dutch blogs by using Dutch version of LIWC. The results of this study also showed that women used more words related to home. In the same vein, Lenard (2016) analysed the speeches of US congressmen and congresswomen and found that congresswomen used more words related to home in comparison to congressmen.

The results of Mann-Whitney U-test conducted for the present study were also consistent with the previous research reported above. This study found that overall women used more words related to home in comparison to men (U = 753112.500; U (standardized) = -11.846; Expected value = 1014444.000; Variance (U) =486690517.083; p-value (Two-tailed) = < 0.0001).

Once the overall tendency in the use of vocabulary was determined by the hypothesis test, further probe was made to discover any subtle similarities or differences in the use of particular expressions. For this purpose, all the words that occurred with 10 or more raw frequency in either corpus were analysed and arranged thematically to see emerging patterns. As shown in Table 5.12, the expressions used by women relate to the internal aspects of home whereas those by men relate to the external ones. Women use words that relate to the themes of accommodation, baking, chores,

furnishing, home occupants, structure, and utensils used at home. Men on the other hand use more words that related to the themes of location, security and legal aspects of home. Since evidence from other societies, as given in the discussion, supports the view that women used more expressions related to home, it is difficult to explain whether this linguistic phenomenon is due to the social role assigned to women or due to women's essential biological and psychological making.

Table 5.12

Gender differences in themes related to 'home'

MenCorB	WenCorB
Location: address	**Accommodation:** apartment, apartments, home, homes, house, houses, housing
Exterior: Backyard, gate	**Baking**: bake, baked, baker, bakeries, bakery, baking
Security: gatekeeper	**Chores/Hobbies:** chores, gardening, homework, household, households, housewife, housewives, housework, maid maids, nannies, nanny, sweeper, sweepers sweeping
Legal: landlord, landlords, lease, leased, mortgaged, rent, rental, tenants	**Interior Furnishing**: bed, beds, carpet, carpets, closet, couch, curtains, furniture rug, sofa, sofas, draped, draper
	Occupants: neighbor/neighbour, resident(s), family, homesick, pet pets, roommate, families
	Structure: balcony, bathroom, bathrooms bedroom, door, doorbell, doors, doorstep doorsteps, dorm, garden, gardens, gardens gates, kitchen, kitchens, laundry, room, rooms
	Utensils: broom, fridge, microwave, oven pillow, shower, stove

Money

Some of the previous research has combined money and work (e.g. Bischoping, 1993) and noticed an increasing trend of money reference in women's speech. Since then, there has been an increasing interest in gender differences with regard to references to use of expressions related to money. Later research conducted in different contexts (Lester, 2004; Nowson, 2006; Schler, Koppel, Argamon, & Pennebaker, 2005; Yale, 2007; Newman, Groom, Handelman, & Pennebaker, 2008; Cunha, Magno, Gonçalves, Cambraia, & Almeida, 2013; Jie, 2015) has consistently reported money to be a theme or feature in men's discourse.

The Mann-Whitney U-test conducted for the present study showed statistically significant difference between men and women regarding the use of words related to money. The test yielded results consistent with the previous studies that had reported overuse of money references in men's texts (U = 1117462.500; U (standardized) = 4.664; Expected value = 1014444.000; Variance (U) = 487968165.251; p-value (Two-tailed) = < 0.0001).

A possible explanation for this tendency could be that for about past one thousand years, the traditional concept that prevails about genders is that of "men outside, women inside", which means that men control the external world and go out to earn money whereas women perform activities at home (Jinyu, 2014). This trend has also been reported by CL research (e.g. Pearce, 2008.), which confirmed this historical and traditional tendency by examining collocational behaviour of MAN and WOMAN in the BNC.

Religion

Religion has always played an important part in society. Even in today's modern societies, majority of people regard themselves as religious, a fact that may surprise someone who is

influenced by the seemingly secular ideology of the modern times. In 2008, Gallup conducted a poll across 143 countries and territories asking participants whether religion was an important part of their lives. About 82% of the respondents answered positively (Janmohamed, 2010). However, one can differentiate between institutionalised religions acting through organisations and an individual's attitude towards faith, which shows his or her religiosity. It is unfortunate that this important aspect of human life, i.e. religion, has been given little attention in language research. As Jule (2005, p.2) says, "It appears religion sits well in the fields of sociology, anthropology and theology, but it is almost unexplored within linguistic research."

Therefore, previous research on gender differences in words referring to religion was even more scarce. Out of the literature explored, only one LIWC-based study by Lenard (2016) could be found that focused on this aspect of research in linguistics. This LIWC-based research was conducted on analysing gender differences in corpora of political speeches delivered by US congressmen and congresswomen. It found that men used more words referring to religion with far more frequency than women. However, the Mann-Whitney U-test conducted for the present study did not find any significant difference between men and women in the use of expressions related to religion (U = 1037167; U (standardized) = 1.030; Expected value = 014444.000; Variance (U) = 486825519.325; p-value (Two-tailed) = 0.303).

After the hypothesis test, the data was further analysed to find out differences in the use of particular words used more by both the genders. For this purpose, words that occurred with a raw frequency of 10 or more in either corpus were subjected to further frequency comparisons. Words used more by men and women were then segregated and grouped under common themes. As shown in Table 5.13, three distinct themes emerged from this analysis. It was found that words

Table 5.13

Gender differences in themes related to 'religion'

MenCorB	WenCorB
Acts of Devotion: morality; pious; pilgrims; rites; ritualistic; meditation; worshipped; faithfully; ritualism	**Acts of Devotion:** prayers; pray; prayer; sacrifice; Sacrifices; Praying; Ramadan; ritual(s); religiously; moral(s); piety; piety; prayed; sacrificing; pilgrimage; prays; worshipping; worshippers
Deity: Krishna; divinely	**Deity:** God; Allah; divine; goddess
Evil: satanic; pagan	**Evil:** Sin; immoral; sins; sinner
Faith: faith; belief; blessing; bless	**Faith:** beliefs; confess; confessed; confession; confessions; merciful
Followers of a Faith: Muslim(s); Hindu(s); Sikh(s) Buddhist(s)	**Followers of a Faith:** Christian(s); Jew(s); Jewish
Metaphysical Entities: spirit; spiritual; souls; devil; Spirits; demons; evangelical; spirituality; immortal; devils; spirited; demonic; immortality; evangelists	**Metaphysical Entities:** soul; angel; angels; Satan; spiritually; angelic
Religious Places: mosque(s); temple(s); shrine; shrines; church(es); seminary; Vatican	**Religious Places:** Mecca; Imambargah; pew; chapel; convent
Proselytising: imam; saint; clergy; saints; Jesus; pope; Christ; imams; preachers; prophecy; prophets; buddha; priest; priests; prophetic; preacher; monks	**Proselytising**: prophet; preach; preaching; pastor; preaches; monk; nun; missionary
Religion General: religious; Islamic; Islam; Islamic; Islamism; Islamises; illumination; religiosity; theological; Christianity; Hinduism; Buddhism; religion; Sikhism; theology; Islamised; Islamise	**Religion General** : religion; religions; Islamism; islamophobia; Islamite; Islamophobic; Judaism; Islamic ally; Islamiyah
Reward and Punishment: paradise; doom; heavens; doomsday; afterlife; immortalised	**Reward and Punishment**. hell; heaven; blessed; mercy; blessings; doomed; salvation; karma
Scriptures: testament; bible	**Scriptures**: Quran/Quranic: scripture(s)
Sectarianism: sectarian; Sunni(s); sects; sect; catholic; protestant; puritan; zen; Catholics	**Sectarianism**. Shiite
Religious Extremism: fundamentalism; fundamentalist; puritanical; fundamentalists; orthodox; Zionist; Zionists; Zion; orthodoxy	**Gender Specific Religious Matters:** veil; veiled; veils
Holy War: jihad; jihadi; mujahideen; jihadis; jihadist; crusade; jihadists; crusaders	

pertaining to two distinct themes of *religious extremism* and *holy war* were used more by men whereas words related to the theme of *gender-specific religious matters* were used more by women. The occurrence of words related to religious extremism and holy war in both the corpora in general and its overuse in MenCorB in particular can be attributed to the prevailing debates on terrorism in the wake of the 9/11 incident. Under the theme *gender-specific religious matters*, women used more terms related to veil (*purdah* in Urdu) for the obvious reason of 'veil' being a religious obligation that exclusively relates to women. Here gender appears to play its role on language. The findings of the present study are expected to be beneficial for future research on religion in linguistics on religion.

Death

There is also a very limited research conducted in linguistics on studying gender differences in expressions related to death. The two LIWC-based studies (Graells-Garrido, Lalmas, & Menczer, 2015; Lenard, 2016) accessed for discussion in this sub-section found that men used these expressions more in comparison to women. The Mann-Whitney U-test conducted for the present study, however, did not find any significant gender difference in the overall frequency of words related to death in MenCorB and WenCorB (U ; 1015451.500; U (standardized) = 0.046; Expected value = 1014444.000; Variance (U) = 480221112.293; p-value (Two-tailed) = 0.963). A possible explanation for equal tendency in men and women for use of words related to death might be the prevailing security situation of Pakistan in the wake of 9/11 incident where the country suffered severe losses in terms of money and blood. The ensuing debates by both the genders in print and electronic media show their equal concern about the situation.

5.7 Informal Language

A large part of our daily life involves interactions with our family, friends, neighbours, colleagues, officials and others. These interactions vary in the way interlocutors use language to express themselves: for instance, in word choice, pronunciation and/or sentence structure (Reid, 1956; Firth, 1968; Halliday, 1978). An important dimension of this linguistic variation is the level of formality (e.g. Labov, 1966; Biber, Conrad, & Reppen, 1998; Heylighen & Dewaele, 2002; Biber & Conrad, 2009). However, while most people can make an intuitive distinction between formal and informal manners of speech (Creber & Giles, 1983; Lahiri & Lu, 2011), it is an ongoing challenge to grasp the exact relation between particular speech situations and the corresponding linguistic characteristics.

Various studies revealed that certain (para)linguistic features occur more in formal than in informal language or *vice versa*. The concept of formality is not as straightforward as one might think, because formality is influenced by many parameters (Berruto, 2010; Heylighen & Dewaele, 2002), such as the setting, the topic of conversation, the modality (written versus spoken language), and the speaker's audience, which potentially have different impacts on speech behaviour.

Gender differences in language formality have been studied by various researchers with contradictory results. For instance, Labov's (1990, p.210) Principle 1 states, "For stable sociolinguistic variables, men use a higher frequency of non-standard forms than women." However, Heylighen and Dewaele (2002) state that language behaviour of men and women changes with the formality of the situation. They reported that men and women differed in informal chat but, in formal essay writing situation, there was no difference found in the formality level of the language of men and women.

The Mann-Whitney U-test conducted for the present study did not find any significant gender differences in the overall use of informal language by men and women bloggers (U = 944801.500; U (standardized) = -3.156; Expected value = 1014444.000; Variance (U) = 487003078.429; p-value (Two-tailed) = 0.002). This tendency may probably be attributed to the genre of blogs, which give more liberty to bloggers to use informal language. According to Nowson (2006), the language used in blog is less constrained by formality. This is further supported by Levy (2009) who states that blogs encourage self-expression, creativity, ownership and community building through informal writing.

5.7.1 Swear Words

"Swearing is the use of any word or phrase that is likely to cause offence when used in middle class polite conversation" (Thelwall, 2008, p. 1). In society today, swear words are used by both men and women. They are prominent features in the vocabulary in context of anger and frustration. However, other areas of emotional states, such as happiness or surprise, can bring about the use of swear words. In addition, swear words can be used for emphasis in any state of mind.

Previous research suggests that gender, age and social class are important factors in the propensity to swear and the type of swearing used. In particular, men seem to swear more than women (Bailey & Timm, 1976) and males swear especially in all-male groups (Coates, 2003; Bayard & Krishnayy, 2001). In fact, some recent research suggests that men use strong swear words more frequently than women, although women use milder swearing more (McEnery, 2006). However, Thelwall (2008) claims on the basis of his personal observations that young women in the UK seem to swear more than men, even with the strongest swear words.

The Mann-Whitney U-test conducted for the present study did not find any statistically significant difference between the genders for the use of swear words (U = 1011997.500; U (standardized) = -0.138; Expected value = 1014444.000; Variance (U) = 313875837.028; p-value (Two-tailed) = 0.890). Moreover, the mean percentage of swear words (i.e. 0.03%) also show that such words are used very scarcely in the texts by bloggers of both the genders. Additionally, a closer examination of the swear words captured by LIWC2015 revealed that certain swearing expressions in English (e.g. *butt*), which are also strong offensive words in social settings like that of Pakistan, were not actually English swearing words in the blog posts but were part of proper nouns in Urdu (names of persons).

As argued by Bilal, Mubashra, Akram, and Shahzada (2013), electronic discourse in Pakistan inherit some of the global features of CMC but much of its content is influenced by its social, cultural and religious factors that are specific to Pakistan. So, bloggers in Pakistan use a language that is similar in many respects to their counterparts in any other part of the world but they also use their language differently in other respects because of peculiar socio-cultural constraints. One of the major socio-cultural influence on CMC in Pakistan is control of swearing language in electronic communication. The phenomenon of swearing is rare on social media in Pakistan in general and in newspaper blogs in particular as these blogs are regularly edited and are seen by a wide readership who have free access to their contents.

5.7.2 Netspeak

Netspeak, a form of the non-standard language, has been used by the public since 1993 (Bacon 1993). With wider application of the Internet in 1999 and early 2000, the term Netspeak became emblematic of online language and included acronyms, abbreviations, emoticons, and

phonetic respellings from the desire of the Internet participants to communicate quickly, who considered a correctly spelled word a sign of the inarticulate and prized an innovative abbreviation above all (Squires, 2010). Netspeak has been defined as "The set of conventions for writing English in e-mail, IRCs, and newsgroups. Netspeak is characterized by acronyms (such as IMHO or ROFL) and clarifying devices such as emotags and emoticons" (Netspeak, 2003). These conventions are also used to compensate for the paralinguistic and nonverbal cues of spoken discourse in the written Internet communication (Crystal, 2006).

In language and gender research, studies encompassing all features of Netspeak could not be found. Nevertheless, some research focused on the use of emoticons as a feature of Netspeak and reported gender differences in their use with contrasting results. For instance, Huffaker and Calvert (2005) conducted a content analysis of 70 blogs and online journals by teenagers. They found that emoticons trended greater in men's texts in comparison to those of the women's. Similar results were reported by Punyanunt-Carter and Hemby (2006) who studied email correspondence between men and women. However, Witmer and Katzman (1997), who examined a data of 3000 messages obtained from public newsgroups and special interest groups, found that women used more emoticons in comparison to men. The results of Lee's (2003) study on Instant Messaging (IM) were also similar. On the other hand, Huaxue and Dechang (2014) are of the opinion that there are not many gender differences in the use of emoticons.

The Mann-Whitney U-test conducted for the present study did not find any statistically significant difference in the use of Netpseak by men and women bloggers (U = 995747.500; U (standardized) = -0.928; Expected value = 1014444.000; Variance (U) = 405948125.223; p-value (Two-tailed) = 0.353). These results might be beneficial for future gender and language research that takes into account multiple features of Netspeak at a time.

5.7.3 Assent

The category of assent words (e.g. *yes*, *yeah*, *okay*, and *alright, indeed*) measures the level of agreement in communication and show a polite communication strategy (Tausczik & Pennebaker, 2010a; Severance, 2012). Previous research on gender differences in words related to assent show that women use more words expressing agreement and politeness than men. For instance, a study on blogs by Schler, Koppel, Argamon, and Pennebaker (2005) reported that female bloggers used more assent expressions than their male counterparts. In addition to reporting similar results, Severance (2012) found that, in male-dominated groups, women expressed more words related to this category in comparison to male bloggers. In the same vein, Bamman, Eisenstein, and Schnoebelen (2014) analysed a corpus of 14,000 Twitter users and found that women were more users of assent words in comparison to men.

The Mann-Whitney U-test conducted for the present study yielded results that were consistent with those of the previous research. The test indicated statistically significant gender differences in the use of assent words. The results showed that women tended to overuse words related to the category of assent with greater frequency than men (U = 933545; U (standardized) = -3.810; Expected value = 1014444.000; Variance (U) = 450962889.154; p-value (Two-tailed) = 0.000). This tendency in women's texts can be attributed to the supportive and polite nature of women's communication as reported consistently by previous research.

5.7.4 Nonfluencies

Nonfluencies (e.g. *er*, *hmm*, *um*) is another dimension of informal language included in LIWC2015 categories. Previous research on gender differences has shown interest in analysing the use of nonfluencies. For instance, analysing the data of six dyads from 60-minute conversation

between two males and two females, Hirschman (1994) observed that women participants used *mm* and *hmm* at a much higher frequency than male participants. Yale (2007) conducted a LIWC-based quantitative content analysis of instant messaging by college students to see gender differences. They observed that females used *ah*, *hmm*, *ugh* and *grr* more frequently than male participants. Similarly, Rao, Yarowsky, Shreevats, and Gupta (2010) gathered a dataset of microblog posts by 1000 participants on Twitter. Their results also indicated the trend of nonfluencies used more among women.

The results of the present study were not consistent with those given by the studies mentioned above. The Mann-Whitney U-test conducted for the present study found no statistically significant difference between the genders in the use of nonfluencies (U = 990601.500; U (standardized) = -1.094; Expected value = 1014444.000; Variance (U) = 474645542.576; p-value (Two-tailed) = 0.274).

5.7.5 Fillers

Fillers or lexical hedges were considered by Lakoff (1975) as one of women's language characteristics and a marker of weakness and tentativeness in women's language in comparison to men (Holmes, 2013). However, later research in gender differences on the use of fillers has brought about mixed results. For instance, some research found that fillers were used more by women (Manjavacas, 2015; Lenard, 2016; Opina, 2017). Others reported contrasting results with men as overusers of fillers (e.g. Mulac & Lundell, 1986; Mehl, Gosling, & Pennebaker, 2006). Yet some studies did not find any gender differences in the use of fillers (e.g. McFadyen, 1996; Hancock & Rubin, 2015).

The results of the present study were consistent with those of McFadyen (1996), and Hancock and Rubin (2015). The Mann-Whitney U-test conducted for the present study did not find any significant differences between men and women bloggers in the use of fillers (U = 987987.500; (U (standardized) = -2.203; Expected value = 1014444.000; Variance (U) = 144214615.370; p-value (Two-tailed) = 0.028).

5.8 Total Punctuation

Research work on historical development of punctuation (Saenger, 1997; Parkes, 2016) trace an evolution in both the types of marks used for punctuation and their function and provides that English punctuation evolved from *rhetorical* (or *correspondence*) *punctuation* to *grammatical* (or *logical*) *punctuation*. While the former was based on classical model of using markers to show where and how long to observe pause while reading aloud, the latter gave readers clues as to the internal structure of a sentence. These trends continued until the evolution of silent reading (Saenger, 1997) and the advent of print culture by the 18th century. Even this transition continues in the modern age, especially after introduction of CMC, with prose incorporating many features of spoken language (Baron & Ling, 2011).

With advances in CMC, a lot of concerns have been voiced by studies with regard to lexical changes in language used on the Internet and SMS (e.g. use of acronyms instead of words, emoticon, and misspellings) and their impact on language degradation as well as associated pedagogical implication (e.g. Thurlow, 2006; Plester & Wood, 2009). However, there is paucity of research on studying punctuation in CMC, especially differences between men and women in its use (Baron & Ling, 2011).

So far gender differences in overall use of punctuation is concerned, previous research suggests that women's language is more punctuated in comparison to men (Corney, Vel, Anderson, & Mohay, 2002; Mukherjee & Liu, 2010; Parkins, 2012). In communication, punctuation marks are not only used for grammatical rules, they also indicate rhythm, pitch, tone, flow and direction. In particular, the repetition of a punctuation mark is a cue that is similar to letter repetitions in CMC. Studies also found that women have the tendency to overuse repeated punctuation marks in CMC (Kalman & Gergle, 2010).

The results of Mann-Whitney U-test conducted for the present study are consistent with those of the previous research. The hypothesis test conducted for this study found statistically significant gender differences in the use of punctuation. The study found that women used more punctuation marks with far more frequency in comparison to men (U = 849797; U (standardized) = -7.452; Expected value = 1014444.000; Variance (U) = 488115291.924; p-value (Two-tailed) = < 0.0001). Since punctuation has traditionally been considered as a feature of standard language, the overuse of overall punctuation marks by women could be attributed to the tendency of women to prefer standard or overtly prestige language as established by the previous studies (e.g. Labov, 2001; Eckert, 1989; Romaine, 2003).

The punctuation category of LIWC2015 is made up of 11 sub-categories, namely: *period*, *comma*, *colons*, *semicolons*, *question marks*, *exclamation marks*, *dashes*, *quotation marks*, *apostrophes*, *parentheses*, and *other punctuation*. Of these, no significant gender differences was found by Mann-Whitney U-test for four sub-categories, i.e. *period*, *colons*, *parentheses*, and *other punctuation*. Therefore, these categories have been excluded and the remaining seven categories are discussed one by one in the succeeding part of this section.

5.8.1 Commas

Despite efforts, previous research on gender differences in the use of comma could not be found. This study found significant gender differences in the use of the mark. The results indicated that women used more commas in comparison to women (U = 931903.500; U (standardized) -3.736; Expected value = 1014444.000; Variance (U) = 488112155.791; p-value (Two-tailed) = 0.000). Therefore, the findings of the present study might be useful for future research in gender and language on the use of this punctuation mark.

5.8.2 Semicolons

Very limited research is available on gender differences in the use of semicolon. Therefore, comparison of the results for the present study were made with those of only two previous studies that could be found. One was conducted by Chiao-Yule (cited in Hamdan & Hamdan, 2013) on a large dataset comprising two corpora: one corpus of 3 million words collected from 47 books, half of which were authored by men and as much by women; the other corpus comprising 4.22 million words collected from 48 blogs. The study found that male authors used more semicolons than female authors. The second study was conducted by Aperoch (2016) on argumentative essays composed by 50 male and female students. The study, in contrast, found that male students scarcely used semicolons to separate clauses. Instead they used coordinating and subordinating conjunctions for the purpose.

The Mann-Whitney U-test conducted for the present study also found statistically significant differences in the use of semicolons by male and female bloggers. The results showed that women tended to overuse semicolon in their blog posts in comparison to men (U = 923247.500; U (standardized) = -4.162; Expected value = 1014444.000; Variance (U) = 480190653.179; p-value

(Two-tailed) = < 0.0001). These results support the findings of Aperoch (2016). It could be concluded from these results that more research is needed to see which gender prefers semicolons as clause separators in a sentence.

5.8.3 Question Marks

Question marks are indicators of questions in a text. Since Robin Lakoff's (1973) article *Language and Women's Place*, a debate has been sparked off in the field of language and gender on many language features including the overuse of question tags and question intonations in declarative statements by women. She considered overuse of question tags in women's language as a sign of their tentative and trivial language style, which, she thought, disqualified them from position of power. Later Fishman's (1983) study in mixed-sex conversation also found that women used more questions in comparison to men. This study, however, found that women asked more questions to facilitate talk and get their male counterpart open up and talk to them. So, women use questions not because they are powerless but because they perform conversational labour as part of their social life.

Some years later, Coates (1996, 1998) carried out an in-depth analysis of features of women's language that included questioning besides others. She observed that women used more questions as a linguistic strategy to establish close and equal social relationship and minimise social distance. She considered it women's strength of conversational and personal skills instead of their weakness. Similarly, James and Clarke (1993), and Suborn (2013) not only found that women used more questions in comparison to men but also attributed this tendency in women's language to their strategy to encourage others to speak, thus aiming rapport-building.

The Mann-Whitney U-test conducted for the present study also found that women used more question marks in comparison to men (U = 862023.500; U (standardized) = -6.914; Expected value = 1014444.000; Variance (U) = 485944376.695; p-value (Two-tailed) = < 0.0001). These results are not only consistent with the previous research but also are internally consistent with those of the *Interrogatives* category mentioned in Section 5.5.6 of this study.

5.8.4 Exclamation Marks

Exclamation marks are indicators of strong feelings (Greenbaum & Nelson, 2002, p. 17) and "great emotive force" (Quirk, Greenbaum, Leech, & Svartvik, 1985, p. 1633) or "strong assertion" (Allen, 1992). The search or exclamation marks in women's and men's language as an indicator of excitability in modern research can be traced to Hiatt's (1977) *The Way Women Write.* Working on the assumption that certain punctuation marks, including exclamation mark, can be used to trace excitability in gendered language, she studied a 2000-word sample selected from books. Her tallies indicated that men used more exclamatory marks in comparison to women.

Building on one another, later research, however, consistently contradicted her results. For instance, Scates (1981) built on Hiatt's (1977) study and analysed many of the same stylistic features as done by Hiatt (1977) in first-year college students' writing. The study found that women used exclamation marks far more frequently than men. Building on the two studies by Hiatt (1977), and Scates (1981), Rubin and Greene (1992) also analysed college students' writing and found that women used exclamation marks with far more frequency than men. Later, with a little variation in the nature of the composition assignment, Winn and Rubin (2001) analysed college students' self-description compositions for excitability based on Hiatt's (1977) markers and noted that women used these markers three times more frequently than men. Consistent results

were also reported by Colley and Todd (2002) for overuse of exclamation markers, especially multiple exclamation markers, by women in comparison to men.

The Mann-Whitney U-test conducted for the present study also found significant gender differences in the use of exclamation marks. The results indicated that women used these marks with far more frequency than men (U = 882401.500; U (standardized) = -6.220; Expected value = 1014444.000; Variance (U) = 450718139.801; p-value (Two-tailed) = < 0.0001). These results are consistent with the studies mentioned above, which found that women overuse exclamation marks in comparison to men. These results are also internally consistent with women's overuse of words related to *negative emotionality* dimension of LIWC2015 as recorded in this study. The overuse of excalamtion marks in women can be attributed to their expression of more emotions and strong feelings as reported by previous research.

5.8.5 Dashes

Despite efforts, previous research in gendered language on the use of dashes could not be found. Therefore, the findings of the present study might be useful for future research in gender and language on the use of this punctuation mark. This study found significant gender differences in the use of the mark. The results indicated that men used more dashes in comparison to women (U = 1120768.500; U (standardized) = 4.813; Expected value = 1014444.000; Variance (U) = 488048414.923; p-value (Two-tailed) = < 0.0001).

5.8.6 Quotation Marks

In previous research, Lakoff (1975) reported direct quotations as one of the features of women's language. Despite efforts, no other research could be found, which focused on the use of

direct quotation to explore gender differences in language use. The present study also confirmed that women used more quotation marks in comparison to men (U = 820750.500; U = (standardized) -8.772; Expected value = 1014444.000; Variance (U) = 487547625.651; p-value (Two-tailed) = < 0.0001). The results of the present study, therefore, are consistent with Lakoff's (1975).

5.8.7 Apostrophes

In language and gender domain, very limited research is available, which has studied gender differences in the use of apostrophe. In recent past, Squires (2007) treated apostrophe as a linguistic variable and examined gender differences in a corpus of Instant Messages (IM). The study found that women used more apostrophes in comparison to men.

The Mann-Whitney U-test conducted for the present study also found statistically significant differences between genders in the use of the apostrophe. The results found that women used more apostrophe in comparison to men (U = 886262; U (standardized) = -5.802; Expected value = 1014444.000; Variance (U) = 488104005.135; p-value (Two-tailed) = < 0.0001). Thus, the results of the present study are consistent with those of Squires (2007).

Compared to other punctuation marks, the use of apostrophe in a text is clear and easily searchable. Therefore, it can easily be quantified. In standard English, apostrophe serves two chief purposes: possession and contractions (Hacker & Sommers, 2012). The tendency in women to use more apostrophe than men may be one indicator of their tendency to prefer overtly prescribed standard form of language, which is a mark of formal or high-status language. This trend in women is internally consistent with the results of this study for *Clout* variable given in Section 5.3.2 of this study.

6. CONCLUSIONS

6.1 Overview of the Study

The overarching purpose of this study was to examine gender differences in language use by men and women bloggers of English e-newspapers in Pakistan. More formally, the thesis was based on the following hypotheses:

H_0: There is no difference in the language use of men and women bloggers of English e-newspapers of Pakistan.

H_1: There is a significant difference in the language use of men and women bloggers of English e-newspapers of Pakistan.

With this assumption, the study identified the research problem and highlighted its significance in the context of Pakistan in *Chapter 1*. The next *Chapter* covered an exploration of the available literature and recorded a series of findings by various studies with respect to the gender and language differences. Particular attention was given to review previous research on language and gender in Pakistan and explore Pakistani blogosphere. *Chapter 3* dealt with methodological issues concerning the present study. Since this analysis was predominantly quantitative in nature, the selection of appropriate statistical procedures and tests applicable as well as their results were given in *Chapter 4*. After these tests, a comparison of findings by the present study with those of the previous research was made for each of the LIWC2015 variables and explanation for any differences found in the gendered language of the blogs was presented in *Chapter 5*. The final Chapter presents the key findings and limitations of the present study as well as making recommendations for future research.

6.2 Key Findings

The study compared two corpora for finding any gender differences in the language use by men and women. For this purpose, the blogs maintained by five English e-newspapers of Pakistan were selected. LIWC2015 version was used for preliminary analysis of the corpora followed by normality tests of the data and hypothesis test through Mann-Whitney U-test, which were conducted with the help of XLSTAT (Addinsoft, 2016). Since LIWC2015 did not allow further probe into data such as Key Word In Context (KWIC) or concordance analysis of the captured words, another software AntConc 3.4.4 (Anthony, 2016) was used for these purposes, wherever required. The study met its specific objectives and had the following key findings.

6.2.1 Methodological Aspects

LIWC was found to be a useful text analysis software for analysing a large corpus of data. With its inherent features of calculating percentages for word frequencies in multiple files at a time, the tool is useful for conducting hypothesis test. Another advantage of the tool is that its pre-established categories prevent the researcher's bias during the analysis stage. Similarly, the independently rated 93 linguistic categories available in the software offer a wide array of language features to researchers from word count to grammatical features to a variety of themes besides punctuation. More importantly, by virtue of being able to capture language features like *netspeak*, *nonfluencies*, *fillers* and *informal language*, the tool was useful for analysing CMC data such as blogs. While LIWC was found suitable for a top-down hypothesis test, the tool does not allow for analysis of concordances or language in context. Moreover, it cannot disambiguate the meaning of words. To overcome these limitations, the tool has to be combined with a concordance software, like AntConc 3.4.4 (Anthony, 2016) in this study. In addition, mere raw numbers obtained from

LIWC results would be useless if they are not subjected to statistical procedures. In case of the present research, LIWC-based results were further analysed with XLSTAT (Addinsoft, 2016). This statistical package was found a useful supplement because of its ease in execution. The package works as a simple add-in feature in MS Excel and, unlike other tools, does not require researchers to transport data from MS Excel worksheet to another statistical package. The combination of the three software was found sufficient for the aim of the present study, which tried to examine, which group of bloggers used a particular language feature more frequently and statistically more significantly.

6.2.2 Linguistic Aspects

While this study found some statistically significant differences in the use of language for some LIWC2015 categories, it found no statistically significant difference on others. Overall, out of 93 language variables, significant gender difference were found for 65 variables. Among these language categories used differently by the genders, men were found to have overused 15 variables in comparison to women, i.e. *analytical thinking*, *word/sentence*, *words > 6 letters*, *articles*, *prepositions*, *comparisons*, *numbers*, *drives*, *achievement*, *power*, *relativity*, *space*, *work*, *money*, and *dashes*.

On the other hand, women overused 50 language categories in comparison to men, i.e. *clout*, *authentic*, *dictionary words*, *total function words*, *total pronouns*, *personal pronouns*, *1st person singular pronouns*, *1st person plural pronouns*, *2nd person pronouns*, *impersonal pronouns*, *auxiliary verbs*, *common adverbs*, *conjunctions*, *negations*, *common verbs*, *interrogatives*, *affective processes*, *negative emotion*, *anxiety*, *sadness*, *social processes*, *family*, *friends*, *female references*, *cognitive processes*, *insight*, *discrepancy*, *tentative*, *certainty*,

differentiation, *perceptual processes*, *see*, *hear*, *feel*, *biological processes*, *body*, *health*, *sexual*, *ingestion*, *affiliation*, *present focus*, *home*, *assent*, *total punctuation*, *comma*, *semicolons*, *question marks*, *exclamation marks*, *quotation marks*, *apostrophes*.

No statistically significant difference was found for the remaining 28 features, i.e. *word count*, *emotional tone*, *3rd person singular*, *3rd person plural*, *common adjectives*, *quantifiers*, *positive emotion*, *anger*, *male references*, *causation*, *reward*, *risk*, *past focus*, *future focus*, *motion*, *time*, *leisure*, *religion*, *death*, *informal language*, *swear words*, *netspeak*, *nonfluencies*, *fillers*, *period*, *colons*, *parentheses*, and *other punctuation*.

When the results of the present research were compared to the findings of the studies cited in *Chapter 5: Discussions* of the present work, some consistent features surfaced, which merit attention. Of the 15 male-dominated language variables in this study, results for men were consistent with the previous research in 06 categories, i.e. overuse of *articles*, *numbers*, *words>6 letters* and expressions that relate to *analytical thinking*, *power*, and *space* (spatial ability). In contrast, out of 50 female-dominated language variables in this study, women were consistent with the previous research in the overuse of 15 language categories, i.e. *total function words*, *total pronouns*, *personal pronouns*, *negations*, *total punctuation*, *question marks*, *exclamation marks*, *quotation marks*, *apostrophe* and words that relate to *clout*, *present focus*, *sense of touch*, *affiliation*, *home*, and *assent*.

By studying concordances and collocations, however, it was also found that certain subtle differences existed between the genders in the use of particular expressions even from within the categories that did not record statistically significant differences (see Section 6.3 *Additional Findings* of this chapter). It could, therefore, be concluded that mere quantitative approach may

not be sufficient to cover all aspects of an analysis and for a more detailed analysis, qualitative studies on each of the language features will be interesting to see.

Besides exploring gender differences in language use, the present study also aimed to verify the claims of the *essentialist* and *social constructionist* theories. The question this study tried to answer was whether stereotypical language differences as predicted by the *essentialist* theory of gender persist in Pakistani context. Because of the role culture plays as a determinant of linguistic style, as claimed by the Tannens' *difference approach*, this study presumed that language use between men and women would be consistent with the *essentialist* theory of gender and language.

In this regard, based on its findings, this study makes two broader but contrasting conclusions. The first conclusion is that some similarities exist in language use by men and women. Results for 28 variables in this study showed no significant differences between the two genders. A possible explanation for this could be that both the genders do not use totally different languages rather they are "drawing on the same linguistic resources" with an inevitable language overlap to be able to communicate with one another (Johnson, 1997, p.11).

The second conclusion is that there are also many significant differences in the use of language by men and women. Overall, the gendered language differed across 65 variables. These differences can, again, be grouped into two types. The first type of gendered language differences found by this study had also been consistently reported by the previous research. So, when results of the present study were compared to those of the previous research, it was observed that language behaviour of both the genders was consistent across 21 variables in all. Such consistency in findings of studies conducted in varied contexts, both in Pakistan and abroad, tend to reinforce the

essentialist view of gender, which relies on the idea of existence of true distinctive *essences* that can be observed directly, empirically, and objectively (DeLamater & Hyde, 1998).

On the other hand, the second type of differences found by this study in the gendered language use were not consistent with those reported by the previous research. In total, results for 44 variables were of this type. While the role of other factors can also not be ruled out, this inconsistency in gendered language across studies conducted in varied contexts in general and those of the present in particular, tend to favour the argument of *social constructionists* who claim that true essences or realties cannot be known objectively; rather, humans always socially engage to create subjective realties (DeLamater & Hyde, 1998). Thus, it seems that both men and women negotiated their language behaviour to adjust to the target context.

6.2.3 Gender Effect on Language

Effect size for the present study was calculated with the help of Cohen's *d,* which shows "the degree of departure from the null hypothesis of the alternate hypothesis" (Cohen, 1988, p. 20). Cohen (1988) classified effect sizes into three categories, namely small (.2), medium (.5) and large (.8). Most of the effect sizes on the language dimensions analysed in this study were in the range of small and between small and medium according to Cohen's (1988) classification. Among female bloggers, *female references* (.63), *dictionary words* (.60), *social processes* (.57), *total pronouns* (.54), *personal pronouns* (.53), *family* (.52), and *biological processes* (.50) scored between medium and large effect sizes. In men bloggers, *analytical thinking* (.52) crossed the threshold of medium effect sizes. The overall summary of effect sizes is given in Figure 6.1.

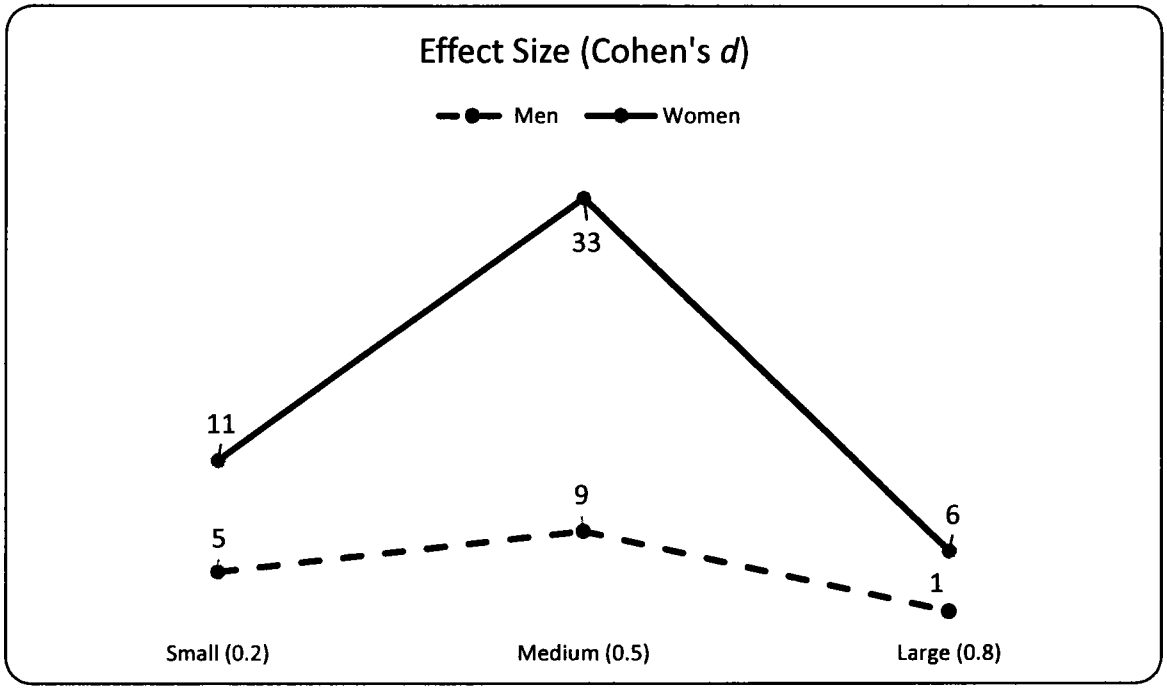


Figure 6.1. Summary of Effect Sizes.

The effect sizes on all language dimensions were in the range generally considered small. In fact, only five dimensions met Cohen's (1988) criterion for small effect—*long words*, *articles*, *swear words*, *social words*, and *pronouns*. These five dimensions appear to be the most meaningful differences in males' and females' language. In interpreting the size of these effects, two considerations are worth noting. These linguistic effect sizes correspond to such trends found in other fields of research on gender differences as synthesised by Eagly (1995) who pointed out that gender differences generally have effect sizes that range from small to medium with relatively unusual large sizes according to Cohen's classification. These results are particularly compelling because of the diverse nature of the bloggers' population, which included language samples from educated members of both the genders that belonged to different walks of life. Despite this, it could be concluded, that men and women used language in reliably and systematically different ways. However, despite these quantitative differences in the use of language, it could not be concluded

that men and women spoke 'different' languages. These differences could rather be seen across a quantitative continuum, which ranges from 'no difference' to 'large size difference'.

6.3 Additional Findings

Besides hypothesis test for the 93 variables, this study also tried to explore some of the subtle language differences and similarities that existed in the texts. As a result of this effort, some interesting additional observations were made. These additional findings shed light on the similarities and differences in language use by men and women pertaining to grammatical and lexical, social, psychological, and biological aspects as well as routines represented in the language of the two genders.

6.3.1 Grammatical & Lexical Aspects

In comparison to women, men used more contracted forms of auxiliaries with all personal pronouns, even if the pronoun referred to female gender (e.g. *she's*). Both men and women used second person pronoun *you* for the 'addressee' and for 'everybody' as theorized by Sacks (1992) and Malone (1997). Additionally, it was found that in the sense of 'everybody', men and women did not always generalize the use of *you* to "any person from any gender" but also used it in the sense of 'every man' or 'every woman', i.e. the gender-directed sense of 'everybody' as shown in Concordances 6.1 and 6.2 below.

cinema in Karachi, Cineplex is restricted to families. So	***you***	can't be a guy and watch a movie in a cinema. No, this

Concordance 6.1 *you* as gender-directed 'everybody': MenCorB

ked what you have studied so far, or what your career plans are.	***You***	may also be asked if you are "aware" of all household chores.
Even though I might not be better than you, I am not any less than	***you***	either just because I cover my face. 7. Don't you feel hot in
picture of you, I was with you. People thought it hypocritical, seeing that	***you***	are an actress who has time and again displayed skin as a

Concordance 6.2 *you* as gender-directed 'everybody': WenCorB

Both the genders used the first-person plural pronoun *we* for identification with some collectivity. In this sense, the pronoun was used by men and women bloggers when they became spokesperson of 'nation', and 'ordinary citizen'. However, men also used *we* for identification when they acted as spokespersons of 'national cricket team' as shown in Concordance 6.3.

strategies are changed. For example, I would understand if	***we***	play only seven batsmen to accommodate a third seamer.
top ranked Pakistani batsman in both ODI and T20 formats.	***We***	are wasting him at number 6 and number 7, he should bat

Concordance 6.3 Spokesperson of 'cricket team': MenCorB

In contrast to men's talking on behalf of the 'national cricket team', women talked as spokespersons of other women, i.e. their own 'sex group' as given in Concordance 6.4.

I'm rambling. Just bear with me. Come on,	***we***	're girls, we should confide in each other. I really like s
er, poverty, disease and exploitation – evils that	***we***	all aspire to see an end to. Not just that, if these young girls

Concordance 6.4 Spokesperson of 'same-sex group': WenCorB

This tendency in females supports Ndambuki and Janks (2010) who concluded that women construct their identities through the members of the gender they belong to.

Similarly, men used *they* in oppositional context for those who violate law, are cruel and corrupt or engage in personal opposition. Thus, men use this context for socio-political or personal purposes. Women, in contrast, used *they* in oppositional context for privacy intruders, violators of women's rights and members of the opposite sex, i.e. men as given in Concordance 6.5.

	the privacy intruder	
ink that child birth is their key to a lock with which	***they***	have now access to the doors of your privacy. They feel
	the violator of women's rights	
her. December 20, 1995 They're beating her again.	***They***	're punishing her for not succumbing to their "pure" de
wheels will always look down upon women drivers.	***They***	wear a disgruntled expression on their face with their ey
	the opposite sex	
Men in their 20s continue to live in a bubble where	***they***	think they will be young forever. But in their early 30s, t

Concordance 6.5 Oppositional context of 'they': WenCorB

Thus, their oppositional context pivots around their own gender-related problems. In this sense, they do not present the oppositional group on behalf of the society but on behalf of their own sex-group. Both men and women overused their own gender-directed third personal singular pronoun (*he*, *she* etc.).

The present analysis also found that the preposition *of* was overused by men while *for* and *with* were overused by women. These results supported findings by Koppel, Argamon and Shimoni (2002). Results of the present study were also consistent with those of the previous research on overuse of colour terms and particular adverbs by women (*awfully*, *vastly*) and men (*more*, *most* and *much*).

6.3.2 Social Aspects

The phrase *her husband* occurred more frequently than *his wife* in GenCorB; women made more frequent references to their husbands than men did to their wives. In contrast, men made more references to their ancestors than women. This reflects the tendency to trace ancestral lineage through male relations in general, particularly in patriarchal societies like that of Pakistan.

Men also used more masculine and less feminine and neuter nouns when talking about relations of other men (e.g. *his father*). Women used more feminine and neuter nouns but less masculine nouns when talking about relations of other men. But, women used more feminine, masculine and neuter nouns when talking about relations of other women. Women used more expressions related to *family*. In particular, they used more words relating to the theme of *marriage,* e.g. *marriage, marriages, wedding, weddings, divorce, honeymoon etc.*

When used with masculine possessive adjective 'his', the relationship of grandfather was used as an ancestor, an authority so be referred to, a model to be emulated, and an ordinary family elder. When used with feminine possessive adjective 'her' the relationship of grandfather was

viewed as an 'ordinary elderly relation'. On the contrary, when used with masculine possessive adjective 'his', 'grandmother' meant an ordinary family elderly relation. When used with feminine possessive adjective 'her', 'grandmother' was viewed as an authority, someone to be taken after, a family elderly relation, and a symbol of tradition. This sheds light on the Pakistani socio-cultural context where men talk more about their male ancestors whereas females do so about their mothers and grandmothers.

Women tended to be concerned with the social status of their own sex-group and sympathize with weak and dependent groups like animals and children. Women also overused words that indicate friendship based on personal association and love. Conversely, men overused words that indicated friendship based on non-personal basis, e.g. business, service, or a common cause.

Women used 'affiliation' words that related to the themes of family, friends, love relation, socialization, communication, social media, affiliative feelings, and common identity. On the contrary, men overused 'affiliation' expressions related to the themes of alliances, brotherhood, grouping, teams, military bodies, socialization, and affiliative feelings. For 'socialization', women used terms related to dance parties and celebrations; men referred to general gatherings.

6.3.3 Psychological Aspects

While expressing positive emotions, women preferred words that convey 'sense of belongingness' such as *love* and *care* whereas men overused words that indicated 'sense of competition, support and achievement', e.g. *play, support, won, win* etc. Similarly, in negative emotions, men overused words that related to 'competition' (e.g. *war, attack*) whereas women overused words that related to 'hatred' (e.g. *hate*) and victimization (such as *rape, victims, abuse*

etc). While expressing anxiety, women overused terms that related to 'internal states' of anxiety, (e.g. *fear guilt, shame*). Men, in contrast, overused terms that relate to some 'external cause' of anxiety (e.g. *pressure, terrorist, threat*). In expression of sadness, women overused terms such as *tears, cry* and *depression* more than men. In affiliative feelings, women used expressions that showed sympathy, compassion, and help whereas men used terms that showed cooperation.

6.3.4 Biological Aspects

With respect to body parts, men overused words that related to physical strength and masculinity (e.g. *beard, muscles, bald*) whereas women overused words that relate to physical beauty and femininity (e.g. *face, skin*). In diseases, men talked more about diabetes; women talked more about other diseases. In negative health habits, men talked more about alcohol, cigar and cigarettes. In food items men talked more about alcohol, soft drinks, and fast food. Since, it is unlawful to consume alcohol as a regular food item in Pakistan, all discussion on alcohol in GenCorB focused on its unlawfulness or adverse social/health effects, which shows the relationship between society and language use.

6.3.5 Routines

In work category, men used words that related to the themes of job, politics, success/failure, business/industry, and leadership/management. In contrast, women tended to overuse words related to the themes of education, health and law. In both the corpora, common themes for leisure activities were travelling, music and entertainment. However, men also overused words related to alcohol, movies/films, literary readings, indoor games, sports/outdoor games and physical fitness whereas women overused words related to art, food and cooking, social media, socialization, rest/relaxation and shopping. Women also tended to overuse expressions related to the internal

aspects of home, e.g. accommodation, baking, chores, furnishing, home occupants, home structure and utensils used at home. On the other hand, men overused words that related to the themes of location, security and legal aspects of home. In the category of religion, men overused words related to the themes of religious extremism and holy war whereas women overused words related to their gender-directed religious matter, e.g. *veil, veiled, veils* for the obvious reason that the injunction of observing 'veil' deals with women. Thus, gender plays its role in language use.

6.4 Strengths and Limitations

In the past, research on gender differences in language use was predominantly conducted in conversation analysis on a case-by-case basis, based on limited data, and relied heavily on manual qualitative analyses by human coders. The works were much time-consuming and painstaking. While these qualitative studies were able to get an in-depth picture of a small amount of data, they were, in no way, designed to provide a linguistic picture or a whole society or culture. This is where quantitative CL research methods, like the present study, can contribute. In particular, the large data available on blogs has been considered suitable for a broader socio-cultural linguistic picture of a society (Pennebaker, 2013). The current study, therefore, bypassed many of the limitations of previous studies and tried to provide a wider understanding of gender differences in language by applying quantitative methods and extending analysis to the realm of a computer-mediated communication text genre, i.e. blogs, analysing a large set of language data, and applying LIWC-based analysis techniques combined with statistical procedures. Thus, the empirical results of the study could be a useful addition to the existing body of literature.

LIWC2015 is a lexical resource containing 6,549 words (Wissen & Boot, 2017), which compares only grapheme patterns of the input files with those incorporated in its default dictionary

(Savović, Ubović, & Radenović, 2018). The tool is based on a crude word-count approach, which does not pay attention to word functions and meanings. However, this is due to the theoretical underpinnings as the software is "instrumental in its aim and thematic in its approach. It captures broad aspects of language use . . . [and] found to be most effective in tracking stylistic aspects of language" (Mehl, 2006, p. 151). Thus, it does not carry out the representational text analysis to look for the sender's intended meaning. For the purpose of the present study, these shortcomings of LIWC2015 were adressed, where more contextuliased analysis of specific words was required, by combining it with AntConc 3.4.4 (Anthony, 2016). LIWC has also been found prone to errors in categorising lexical items. However, with large size data, as is the case with the present research, "the liekly error rate is extremely low" (Newman, Groom, Handelman, & Pennebaker, 2008, p. 217).

This study also limited itself to identification of how men and women use language differently without discussing the impact of age, level of education, genre or any other variable on language. The reason is that language is a complex social phenomenon and the present study could focus only on relationship of language with only one variable, i.e. *gender*, due to limited time and resources. Similarly, bilingual aspects like code-mixing could also not be studied because the ATA tools used in this study could capture English expressions only. Despite these shortcomings, the present study could be considered as a useful beginning in corpus-based empirical studies in Pakistan to trigger debate on gender differences in language use.

6.5 Recommendations for Future Research

The present study also suggests some new ways to explore gender differences in language use by future studies. For instance, one possible direction could be to examine gendered texts from

other web genres like Tweets, and Facebook postings. Similarly, the present study predominantly followed a top-down approach and further research can use bottom-up approaches to analysis of language by men and women. In the same vein, a closer and detailed examination on limited language features from LIWC2015 could be conducted by future studies through mix-method approach. Future research can also conduct a comparison of topic choice by the two genders. The enormous diversity of languages in Pakistan also offers rich opportunities for future studies to explore the relationship between language and gender in bilingual context as well as other unexplored vernaculars. In addition, further expansion in corpus linguistic tools in Pakistan can be made by developing an Urdu dictionary, which can be integrated with LIWC for CL research in the national language of Pakistan.

References

Abbar, S., Mejova, Y., & Weber, I. (2015). You tweet what you eat: Studying food consumption through Twitter. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3197-3206). ACM.

Abdurahman, N. H. (2017). An investigation of different language choice through personal pronouns in the Twitter. *Indonesian Journal of EFL and Linguistics, 2*(1), 1-13. Retrieved January 24, 2018, from http://www.indonesian-efl-journal.org/index.php/ijefll/article/view/18/pdf

Addinsoft (2016). *XLSTAT 2016: Data Analysis and Statistical Solution for Microsoft Excel* [Computer Software]. Paris, France: Addinsoft.

Addis, M. E., & Mahalik, J. R. (2003). Men, masculinity, and the contexts of help seeking. *American Psychologist, 58*(1), 5-14. doi:http://dx.doi.org/10.1037/0003-066X.58.1.5

Adler, E. (2013). *Sex differences in positive emotions within Appraisal Theory* (Doctoral dissertation, Vanderbilt University).

Agarwal, S., Godbole, S., & Punjani, D. (2007). How much noise is too much: A study in Automatic Text Classification. In N. Ramakrishnan, O. R. Zaïane, Y. Shi, C. W. Clifton, & X. Wu (Ed.), *Seventh IEEE International Conference on Data Mining (ICDM 2007) 28-31 Oct. 2007* (pp. 3-12). Omaha, NE, USA: IEEE Computer Society. Retrieved November 15, 2017, from http:// ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=4470209

Ahmad, I., & Mehmood, A. (2015). Gender differences in the use of parts of speech in Pakistani English newspaper columns: A critical discourse analysis. *New Media and Mass Communication, 36*, 10-15. Retrieved January 17, 2018, from http://www. iiste.org/Journals/index.php/NMMC/article/view/21762/21790

Ahmed, K., Ali, I., & Xianfa, P. D. (2015). Code-mixing as a marker of gender identity in SMS language in Pakistan. *IOSR Journal of Humanities and Social Science, 20*(1), 58-65. Retrieved November 4, 2017, from http://www.iosrjournals.org/iosr-jhss/papers/Vol20-issue1/Version-5/H020155865.pdf

Ali, G., & Khan, L. A. (2012). Language and construction of gender: A feminist critique of SMS discourse. *British Journal of Arts and Social Sciences, 4*(2), 342-359. Retrieved November 5, 2017, from http://www.bjournal.co.uk/paper/bjass_4_2/bjass_04_02_17.pdf

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. Retrieved November 4, 2017, from http://pu.edu.pk/images/journal/JEE/PDF-Files/Article%20No.%204_V22_2_12.pdf

Ali, R., & Batool, S. (2015). Stereotypical identities: Discourse analysis of media images of women in Pakistan. *Multidisciplinary Journal of Gender Studies, 4*(1), 690-717. doi:10.4471/generos.2015.54

Allen, A. M., Scheurmann, T. S., Nollen, N., Hatsukami, D., & Ahluwalia, J. S. (2016). Gender differences in smoking behavior and dependence motives among daily and nondaily smokers. *Nicotine & Tobacco Research, 6*(1), 1408–1413.

Allen, R. E. (1992). Exclamation mark. In T. McArthur (Ed.), *The Oxford companion to the English language* (p. 394). Oxford: Oxford University Press.

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.

Amjad, H., & Rasul, S. (2017). Female gender portrayal through news blogs: Analysing the ideological representation of Pakistani working women. *FWU Journal of Social Sciences, 11*(1), 163-177. Retrieved October 30, 2017, from https://search.proquest.com/docview/1929032881/fulltextPDF/4E1B0D5A2ABE4341PQ/1?accountid=8630

Ancestry. (n.d.). *http://www.ancestry.co.uk.* Retrieved from Tips for Finding the Stories of the Women in Your Family Tree: http://c.ancestry.com/cs/media/finding-women-stories.pdf

Anderson, T., & Darling, D. (1954). A test of goodness of fit. *Journal of the American*, 765-769.

Andersson, L. (2012). *Personal pronouns in editor's letters – A gender-based study.* Linnaeus University, School of Language and Literature. Växjö, Sweden: Linnaeus University. Retrieved January 14, 2018, from https://www.diva-portal.org/smash/get/diva2:647390/FULLTEXT01.pdf

Anthony, L. (2013). A critical look at software tools in corpus linguistics. *Linguistic Research, 30*(2), 141-161. Retrieved November 24, 2017, from http://www.laurenceanthony.net/research/20130827_linguistic_research_paper/linguistic_research_paper_final.pdf

Anthony, L. (2016). *AntConc (Version 3.4.4)* [Computer Software]. Tokyo, Japan: Waseda University. Retrieved November 23, 2017, from http://www.laurenceanthony.net/

Aperoch, M. D. (2016). Lexical and syntactic features of the male and female students argumentative essays. *UMIMRAJ, 1*(1), 213-226.

Ardila, A., Rosselli, M., Matute, E., & Inozemtseva, O. (2011). Gender differences in cognitive development. *Developmental Psychology, 47*(4), 984-490.

Argamon, S., Koppel, M., Fine, J., & Shimoni, A. R. (2003). Gender, genre, and writing style in formal written texts. *Text - Interdisciplinary Journal for the Study of Discourse, 23*(3), 321-346. doi:https://doi.org/10.1515/text.2003.014

Argamon, S., Koppel, M., Pennebaker, W., & Schler, J. (2007). Mining the blogosphere: Age, gender and the varieties of self-expression. *First Monday, 12*(9). Retrieved January 11, 2018, from http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2003

Aries, E. (1976). Interaction patterns and themes of male, female, and mixed groups. *Small Group Behavior, 7*(1), 7-18. Retrieved Obtober 19, 2017, from http://journals.sagepub.com/doi/abs/10.1177/ 104649647600700102

Aries, E. (1996). *Men and women in interaction.* Oxford, UK: Oxford University Press.

Aries, E. J. (1982). Verbal and nonverbal behavior in single-sex and mix-sex groups. *Psychological Reports, 51*(1), 127-134. Retrieved October 22, 2017, from http://journals.sagepub.com/doi/abs/10.2466/pr0.1982.51.1.127

Aries, J., & Johnson, F. L. (1983). Close friendship in adulthood: Conversational content between same-sex friends. *Sex Roles, 9*(12), 1183-1196. doi:https://doi.org/10.1007/BF00303101

Armstrong, C. L., Wood, M. L., & Nelson, M. R. (2006). Female news professionals in local and national broadcast news during the buildup to the Iraq war. *Journal of Broadcasting & Electronic Media, 50*(1). doi: 10.1207/s15506878jobem5001_5

Ashmore, R. D., & Del Boca, F. K. (1979). Sex stereotypes and implicit personality theory: Toward a cognitive-social psychological conceptualization. *Sex Roles, 5*(2), 219-248. Retrieved January 10, 2018, from https://doi.org/10.1007/BF00287932

Ashmore, R. D., Del Boca, F. K., & Wohlers, A. J. (1986). Gender stereotypes. In R. Ashmore, & F. D. Boca (Eds.), *The social psychology of female-male relations.* Orlando: FL: Academic.

Aston, G. (2001). Text categories and corpus users: A response to David Lee. *Language Learning & Technology, 5*(3), 73-76. Retrieved November 20, 2017, from http://llt.msu.edu/vol5num3/ pdf/aston.pdf

Aston, G., & Burnard, L. (1998). *The BNC handbook.* Edinburgh, UK: Edinburgh University Press.

Azher, M., Khalid, S., & Mehmood, R. (2014). Gender representation in Pakistani and British English fiction: A corpus based study. *Research on Humanities and Social Sciences, 4*(14), 1–6.

Aziz, M., & Rafi, M. (2010). Sentence based semantic similarity measure for blog-posts. In Y. N. Franz Ko (Ed.), *6th International Conference on Digital Content, Multimedia Technology and its Applications* (pp. 69-74). Seoul, Korea: IEEE. Retrieved November 6, 2017, from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5568598

Babayigit, S. (2015). The dimensions of written expression: Language group and gender differences. *Learning and Instruction, 35*, 33-41. doi:https://doi.org/10.1016/j.learninstruc.2014.08.006

Bacon, I. (1993). Out on the internet; bridges; includes related articles on inexpensive internet tools internet jargon, how to get access. *MacUser, 9*, 209–14.

Bailey, L. A., & Timm, L. (1976). More on women's — and men's — expletives. *Anthropological Linguistics, 18*(9), 438-449.

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

Baker, P. (2010a). *Sociolinguistics and corpus linguistics.* Edinburgh, UK: Edinburgh University Press Ltd.

Baker, P. (2010b). Will Ms ever be as frequent as Mr? A corpus-based comparison of gendered terms across four diachronic corpora of British English. *Gender and Language, 4*(1), 125–149. doi: 10.1558/genl.v4i1.125

Baker, P. (2011). Social involvement in corpus studies. In V. Viana, S. Zyngier, & G. Barnbrook (Eds.), *Perspectives on corpus linguistics* (pp. 17-27). Amsterdam, Netherlands: John Benjamins Publishing Company.

Baker, P. (2014). *Using corpora to analyse gender.* London: Bloomsbury Academic.

Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics.* Edinburgh, UK: Edinburgh University Press Ltd.

Bales, R. F. (1950). *Interaction process analysis: A method for the study of small groups.* Reading, MA: Addison-Wesley.

Bamman, D., Eisenstein, J., & Schnoebelen, T. (2014). Gender identity and lexical variation in social media. *Journal of Sociolinguistics, 18*(2), 135-160. Retrieved from http://onlinelibrary.wiley.com.ezproxyd.bham.ac.uk/doi/10.1111/josl.12080/pdf

Banks, I. (2001). No man's land: men, illness and the NHS. *British Medical Journal, 323*, 1058–1060. Retrieved from doi: https://doi.org/10.1136/bmj.323.7320.1058

Bantum, E. O., & Owen, J. E. (2009). Evaluating the validity of computerized content analysis programs for identification of emotional expression in cancer narrative. *Psychological Assessment, 21*, 79-88.

Baron, N. S. (2004). See you online: Gender issues in college student use of instant messaging. *Journal of Language and Social Psychology, 23*, 397-423. Retrieved January 25, 2018, from https://doi.org/10.1177/0261927X04269585

Baron, N. S., & Ling, R. (2011). Necessary smileys & useless periods: Redefining punctuation in electronically-medicated communication. *Visible Language, 45*(1/2), 45-67.

Basow, S. A., & Rubenfeld, K. (2003). "Troubles talk": Effects of gender and gender typing. *Sex Roles: A Journal of Research, 48*, 183-187.

Baumann, A., Krasnova, H., Veltri, N., & Yusni, Y. (2015). Men, women, microblogging: Where do we stand? *Proceedings of 12th International Conference on Wirtschaftsinformatik* (pp. 857-871). Osnabrück, Germany: Universität Osnabrück, Deutschland.

Bayard, D., & Krishnayy, S. (2001). Gender, expletive use, and context: Male and female expletive use in structured and unstructured conversation among New Zealand university students. *Women and Language, 24*(1), 1-15.

Beamish, R. (1998). The local newspaper in the age of multimedia. In B. F. Murphy (Ed.), *Making local news: Local journalism in context* (pp. 140-153). London, UK: Routledge.

Beeching, K. (2002). *Gender, politeness and pragmatic particles in French.* Amsterdam, The Netherlands: John Benjamins Publishing Company.

Beißwenger, M., & Storrer, A. (2008). Corpora of computer-mediated communication. In A. Lüdeling, & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (pp. 292-308). Berlin, Germany: Walter de Gruyter.

Bell, C., McCarthy, P., & McNamara, D. (2006). Variations in language use across gender: Biological versus sociological theories. In R. Sun, & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1009-1013). Mahwah, New Jersey, USA: Erlbaum.

Bennett, G. R. (2010). *Using corpora in the language learning classroom.* Ann Arbor: University of Michigan Press.

Bergh, G. (2005). Min(d)ing English language data on the web. What can Google tell us? *International Computer Archive of Modern and Medieval English (ICAME), 29*, 25–46. Retrieved October 24, 2017, from http://clu.uni.no/icame/ij29/

Bergh, G., & Zanchetta, E. (2008). Web linguistics. In A. Lüdeling, & M. Kytö (Eds.) *Corpus linguistics: An international handbook* (pp. 309-317). Berlin, Germany: Walter de Gruyter GmbH & Co.

Bergvall, V.L. (1999). Toward a comprehensive theory of language and gender. *Language in Society, 28*, 273-293.

Berlin, B., & Kay, P. (1969). *Basic color terms.* Berkeley: University of California Press.

Berruto, G. (2010). Identifying dimensions of linguistic variation in a language space. In P. Auer, & J. Schmidt (Eds.), *Language and space: An international handbook of linguistic variation* (Vol. 1, pp. 226–241). Berlin: Walter de Gruyter.

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

Biber, D., & Finegan, E. (1989). Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text, 9*(1), 93-124.

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing, 4*(4), 243-257. Retrieved November 16, 2017, from https://academic.oup.com/dsh/article abstract/8/4/243/928942

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

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

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

Bilal, H. A., Mubashra, S., Akram, F., & Shahzada, A. (2013). A sociolinguistic study of varying level of formality of computer mediated communication in Pakistan. *International Journal of Humanities and Social Science, 3*(13), 291-297.

Bischoping, K. (1993). Gender differences in conversation topics. *Sex Roles, 28*(1), 1-18. doi:https://doi.org/10.1007/BF00289744

Bishop, D., & Witt, P. (1970). Sources of behavioral variance during leisure time. *Journal of Personality and Social Psychology, 16*, 352-360.

Blader, S. L., & Chen, Y. R. (2012). Differentiating the effects of status and power: A justice perspective. *Journal of Personality and Social Psychology, 102*, 994-1014.

Blei, D., Ng, A., & Jordan, M. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*(4/5), 993–1022. Retrieved October 23, 2017, from http://web.a.ebscohost. com/ehost/pdfviewer/pdfviewer?vid=1&sid=e79b38f1-81f1-453a-99e9-3a29d769b727% 40sessionmgr4008

Blood, R. (2004). How blogging software reshapes the online community. *Communication of the ACM, 47*(12), 53-55. doi:10.1145/1035134.1035165

Bloome, D., & Green, J. (2002). Directions in the sociolinguistic study of reading. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 395-421). Mahwah, NJ: Lawrence Erlbaum.

Boehm, C. (1999). *Hierarchy in the forest: The evolution of egalitarian behavior.* Boston, MA, USA: Harvard University Press.

Bond, G. D., & Lee, A. Y. (2005). Language of lies in prison. *Applied Cognitive Psychology, 19,* 313-329. doi:10.1002/acp.1087

Bouffard, L., Bastin, E., & Lapierre, S. (1994). *The personal future in old age.* Lublin, Poland: Lublin University Press.

Bradley, P. (1981). The folk-linguistics of women's speech: An empirical examination. *Communication Monographs, 48,* 73-90.

Bresnan, J., & Spencer, J. (2013). *Frequency and variation in English subject-verb contraction.* Stanford, CA (USA): Stanford University. Retrieved January 24, 2018, from http://web.stanford.edu/~bresnan/contraction.pdf

Brezina, V. (2018). *Statistics in corpus linguistics.* Cambridge: Cambridge University Press. doi: https://doi.org/10.1017/9781316410899

Brody, L., & Hall, J. (2000). Gender, emotion, and expression. In M. Lewis, & J. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 395-408). New York: Guilford.

Brönnimann, R., Herlihy, J., Müller, J., & Ehlert, U. (2013). Do testimonies of traumatic events differ depending on the interviewer? *The European Journal of Psychology Applied to Legal Context, 5*(1), 97-121.

Brouwer, D., Gerritsen, M., & De Haan, D. (1979). Speech differences between women and men: On the wrong track? *Language in Society, 8*(1), 33-50.

Broverman, I., Vogel, S., Broverman, D., Clarkson, F., & Rosenkrantz, P. (1972). Sex role stereotypes: A current appraisal. *Journal of Social Issues, 28,* 59-92.

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 4,* 822-848.

Brownlow, S., Rosamond, J., & Parker, J. (2003). Gender-linked linguistic behavior in television interviews. *Sex Roles, 49*(3), 121-132. doi:https://link-springer-com.ezproxyd.bham.ac.uk/content/pdf/10.1023%2FA%3A1024404812972.pdf

Bryant, F. B. (2003). Savoring beliefs inventory (SBI): A scale for measuring beliefs about savouring. *Journal of Mental Health, 12*(2), 175-196.

Bureau of Statistics. (2017). *Press release on provisional summary results of 6th population and housing cesnsus-2017.* Ministry of Statistics, Statistics Division. Islamabad: Government

of Pakistan. Retrieved November 1, 2017, from http://www.pbs.gov.pk: http://www.pbs.gov.pk/sites/default/files//press_releases/2017/Population_Census_2017_Results.pdf

Butler, J., & Trouble, G. (1990). *Feminism and the subversion of identity*. New York: Routledge.

Butler, J. (1993). *Bodies that matter: On the discursive limits of "sex"*. London: Routledge.

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin, 3*, 367-383.

Caldas-Coulthard, C. R., & Moon, R. (2010). 'Curvy, hunky, kinky': Using corpora as tools for critical analysis. *Discourse & Society, 21*(2), 99-133. doi:10.1177/0957926509353843

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

Cameron, D. (2009). Sex/gender, language and the new biologism. *Applied Linguistics, 31*(2), 1–20. doi:doi:10.1093/applin/amp022

Cameron, D. (2007). *The myth of Mars and Venus*. London: Oxford University Press.

Cameron, D. (2005). Language, gender and sexuality: Current issues and new directions. *Applied Linguistics, 26*(4), 682-502.

Cameron, D., & Kulick, D. (2003). *Language and sexuality*. Cambridge: Cambridge University Press.

Cammarota, G., M.P., G., Laurino, C., & Pellicano, M. (2008). Gender differences in the eating behaviour of Italian schoolchildren. *La Rivista di Scienza dell'Alimentazione, 2*, n.p.

Carlson, D. (2009). *David Carlson's virtual world*. Retrieved October 28, 2017, from www.iml.jou.ufl.edu: http://iml.jou.ufl.edu/carlson/1990s.shtml#1994

Carli, L. L. (1990). Gender, language, and influence. *Journal of Personality and Social Psychology, 59*(5), 941-951.

Casey, M. B., Nutall, R., Pezaris, E., & Benbow, C. P. (1995). The influence of spatial ability on gender differences in math college entrance test scores across diverse samples. *Developmental Psychology, 31*, 697-705.

Cassell, J., Huffaker, D., Tversky, D., & Ferriman, K. (2006). The language of on-line leadership: Gender and youth engagement on the internet. *Developmental Psychology, 42*(3), 436-449. Retrieved January 9, 2018, from http://citeseerx.ist.psu.edu/viewdoc/download? doi=10.1.1.550.3195&rep=rep1&type=pdf

Catak, P. D. (2012). The Turkish version of the cognitive and affective mindfulness scale-revised. *Europe's Journal of Psychology, 8*(4), 603-619.

Chafe, W. (1995). Polyphonic topic development. In T. Givón (Ed.), *Conversation: Cognitive, communicative and social perspectives* (pp. 41-55). Amsterdam: John Benjamins.

Charteris-Black, J. (2009). Metaphor and gender in British parliamentary debates. In K. Ahrens (Ed.), *Politics, gender and conceptual metaphors* (pp. 139-165). Hampshire, UK: Palgrave Macmillan.

Chaseling, J., Steele, M., Smart, N., & Hurst, C. (2012). A comparison of the power of the discrete Kolmorgorov-Smirnov and Chi-Square goodness of fit tests. In K. Kumar, & A. Chaturvedi (Eds.), *Some recent developments in statistical theory and applications* (pp. 104-109). Boca Raton, Florida, USA: Brown Walker Press.

Chentsova-Dutton, Y., & Tsai, J. (2007). Gender differences in emotional response among European Americans and Hmong Americans. *Cognition and Emotion, 21*(1), 162-181.

Chraif, M., & Anitei, M. (2013). Gender differences in measuring positive and negative emotions self-perception among Romanian high school students–A pilot study. *Procedia - Social and Behavioral Sciences, 76*(15), 181–185.

Chung, C. K., & Pennebaker, J. W. (2007). The psychological function of function words. In K. Fielder (Ed.), *Frontiers of social psychology: Social communication* (pp. 343-359). New York: Psychological Press.

Clancy, B. (2010). Building a corpus to represent a variety of a language. In A. O'Keeffe, & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 80-92). Routledge.

Coates, J. (1993). *Women, men and language.* London: Longman.

Coates, J. (1996). *Women talk: Conversation between women friend.* Oxford: Blackwell.

Coates, J. (1997). The construction of a collaborative floor in women's friendly talk. In T. Givón (Ed.), *Conversation: Cognitive, communicative and social perspectives* (pp. 55-89). Philadelphia: John Benjamins.

Coates, J. (1998). *Language and gender: A reader.* London: Blackwell.

Coates, J. (2003). *Men talk.* Oxford: Blackwell.

Coates, J. (2004). *Women, men and language* (3rd ed.). Harlow: Longman.

Coates, J. (2007). Gender. In C. Llamas, L. Mullany, & P. Stockwell (Eds.), *The Routledge companion to sociolinguistics* (pp. 62-68). Oxford: Routledge.

Coates, L., & Johnson, T. (2001). Towards a social theory of gender. In W. P. Robinson & H. Giles (Eds.), *The new handbook of language and social psychology* (pp. 345–398). New York: Wiley.

Coates, J., & Jordan, M. (1997). Que(e)rying friendship: Discourses of resistance and the construction of gender subjectivity. In A. Livia, & K. Hall (Eds.), *Queerly phrased: Language, gender and sexuality.* (pp. 214-232). New York: Oxford University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Lawrence Erlbaum Associates, Publishers.

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science, 15*(10), 687-693.

Colley, A., & Todd, Z. (2002). Gender linked differences in the style and content of Emails to friend. *Journal of Language and Social Psychology, 4*, 380-392. doi:10.1177/ 0261927022 37955

Corney, M. W. (2003). *Analysing e-mail text authorship for forensic purposes.* Queensland: Queensland University of Technology, Queensland. Retrieved January 12, 2018, from https://eprints.qut.edu.au/16069/1/Malcolm_Corney_Thesis.pdf

Corney, M., Vel, O. d., Anderson, A., & Mohay, G. (2002). Gender-preferential text mining of e-mail discourse. *18th annual Computer Security Applications Conference (ACSAC), 9-13 Dec. 2002* (pp. 282-289). Las Vegas, NV, USA: IEEE Computer Society Washington, DC, USA. Retrieved from https://dl.acm.org/citation.cfm?id=784775

Corver, N., & Riemsdijk, H. v. (2001). Semi-lexical categories. In N. Corver, & H. v. Riemsdijk (Eds.), *Semi-lexical categories: The function of content words and the content of function words* (pp. 23-66). Berlin: Mouton de Gruyter.

Cramer, D., & Howitt, D. (2004). *The SAGE dictionary of statistics: A practical resource for students in the social sciences.* London, UK: SAGE Publications Ltd.

Crawford, M. (1995). *Talking difference: On gender and language.* London, UK: Sage Publications.

Creber, C., & Giles, H. (1983). Social context and language attitudes: The role of formality-informality of the setting. *Language Sciences, 5*(2), 155–161.

Crosby, F., & Nyquist, L. (1977). The female register: An empirical study of Lakoff's hypotheses. *Language in Society, 6*(3), 313-322.

Crystal, D. (2006). *Language and the internet.* Cambridge: Cambridge University Press.

D'Agostino, R. S. (1986). *Goodness-of-fit techniques.* New York: Marcel Dekker, Inc.

Dabs, J. M., Chang, E. L., Strong, R. A., & Milun, R. (1998). Spatial ability, navigation strategy, and geographic knowledge among men and women. *Evolution and Human Behavior, 19*, 89–98.

Danet, B. (1998). Text as mask: Gender, play and performance on the Internet. In S. G. Jones (Ed.), *Cybersociety 2.0: Computer-mediated communication and community revisited* (pp. 129–158). Thousand Oaks, California, USA: Sage.

Danner, D. D., Snowdon, D. A., & Friesen, W. V. (2001). Positive emotions in early life and longevity: Findings from the nun study. *Journal of Personality & Social Psychology, 80*, 804-813.

Das, K. R., & Imon, A. H. (2016). A brief review of tests for normality. *American Journal of Theoretical and Applied Statistics, 5*(1), 5-12. Retrieved January 6, 2018, from http://www.sciencepublishinggroup.com/j/ajtas

Dekeyser, M., Raes, F., Leijssen, M., Leysen, S., & Dewulf, D. (2008). Mindfulness skills and interpersonal behaviour. *Personality and Individual Differences, 44*(5), 1235-1245.

DeLamater, J. D., & Hyde, J. S. (1998). Essentialism vs. social constructionism in the study of human sexuality. *Journal of Sex Research, 35*(1), 10-18. doi: http://dx.doi.org/10.1080/00224499809551913

Delgado, A. R., & Prieto, G. (1996). Sex differences in visuospatial ability. Do performance factors play such an important role? *Memory and Cognition, 2*(4), 504–510.

Depner, C., & Verbrugge, L. (1980). Social networks and health behavior. In M. D. (Ed.), *Women's lives: New theory, research and policy* (pp. 94-128). University of Michigan.

DeYoung, C., Weisberg, Y., Quilty, L., & Peterson, J. (2013). Unifying the aspects of the big five, the interpersonal circumplex, and trait affiliation. *Journal of Personality, 81*(5), 465–475.

Dilshad, M., Adnan, A., & Akram, A. (2013). Gender differences in reading habits of university students: An evidence from Pakistan. *Pakistan Journal of Social Sciences (PJSS)*, 311-320.

Djikic, M., Oatley, K., & Peterson, J. B. (2006). The bitter-sweet labor of emoting: The linguistic comparison of writers and physicists. *Creativity Research Journal, 18*(2), 191–197. doi:10.1207/s15326934crj1802_5

Dovidio, J. F., Brown, C. E., Heltman, K., Ellyson, S. L., & Keating, C. F. (1988). Power displays between women and men in discussions of gender-linked tasks: A multichannel study. *Journal of Personality & Social Psychology, 55*, 580–587.

Driscoll, I., Hamilton, D., Yeo, R., Brooks, W., & Sutherland, R. (2005). Virtual navigation in humans: The impact of age, sex, and hormones on place learning. *Hormones and Behavior, 47*, 326-235.

Dubois, B. L., & Crouch, I. (1975). The question of tag questions in women's speech: They don't really use more of them, do they? *Language in Society, 4*(3), 289–294.

Eagly, A. H. (1987). *Sex differences in social behavior: A social role interpretation.* Hillsdale, NJ: Lawrence Erlbaum.

Eagly, A. H. (1995). The science and politics of comparing women and men. *American Psychologist, 50*(3), 145-158.

Eckert, P. (1989). The whole woman: Sex and gender differences in variation. *Language and Variation and Change, 1*(1), 245-267.

Eckert, P., & McConnell-Ginet, S. (1992a). Communities of practice: Where language, gender and power all live. In K. Hall, M. Bucholtz, & B. Moonwomon (Ed.), *Locating power: Proceedings of the second Berkeley women and language conference* (pp. 89–99). Berkeley, CA: Berkeley Women and Language Group.

Eckert, P., & McConnell-Ginet, S. (1992b). Think practically and look locally: Language. *Annual Review of Anthropology, 21*, 461–490. Retrieved October 6, 2017, from http://www.annualreviews.org/doi/pdf/10.1146/annurev.an.21.100192.002333

Edelsky, C. (1981). Who's got the floor? *Language in Society, 10*(3), 383-421. Retrieved January 7, 2018, from https://www.cambridge.org/core

Eggins, S., & Slade, D. (1997). *Analysing casual conversation.* London: Cassell.

Ek, S. (2013). Gender differences in health information behaviour: a Finnish population-based survey. *Health Promotion International, 30*(3), 736-745. doi:doi:10.1093/heapro/dat063

Epstein, E., Sloan, D., & Marx, B. (2005). Getting to the heart of the matter: Written disclosure, gender, and heart rate. *Psychosomatic Medicine, 67*, 413–419.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49*, 709–724. Retrieved January 8, 2018, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.466.600&rep=rep1&type=pdf

Epstein, S. (2003). Cognitive-experiential self-theory of personality. In T. Millon, & M. J. Lerner (Eds.), *Handbook of psychology: Vol. 5, personality and social psychology* (pp. 159–184). Hoboken, NJ: Wiley. Retrieved January 8, 2018, from http://www.communicationcache.com/uploads/1/0/8/8/10887248/handbook_of_psychology_volume_5_personality_and_social_psychology_2003.pdf

Erickson, B., Lind, A. E., Johnson, B., & O 'Barr, W. (1978). Speech style and impression formation in a court setting: The effects of "powerful" and "powerless" speech. *Journal of Experimental Social Psychology, 14*, 266-279.

Eveland, W. P., & Dylko, I. (2007). Reading political blogs during the 2004 election campaign: Correlates and political consequences. In M. Tremayne (Ed.), *Blogging, citizenship and the future of media* (pp. 105-126). New York, USA: Routledge.

Everett, C. E. (2011). Transformation of newspapers in the technology era. *The Elon Journal of Undergraduate Research in Communications, 2*(2), 102-114. Retrieved October 28, 2017, from https://www.elon.edu/u/academics/communications/journal/wp-content/uploads/sites/153/2017/06/ 09EverettEJFall11.pdf

Fabes, R., & Martin, C. (1991). Gender and age stereotypes of emotionality. *Personality and Social Psychology Bulletin, 17*, 532-540.

Falk, J. (1980). The conversational duet. *Proceedings of the Sixth Annual Meeting of the Berkeley Linguistics Society* (pp. 507-514). California: Berkeley Linguistics Society. Retrieved January 7, 2018, from http://linguistics.berkeley.edu/bls

Farhi, P. (2009). *Search archive.* Retrieved October 28, 2017, from www.ajrarchive.org: http:// www. ajrarchive.org/article.asp?id=4772

Farrel, P., & Stewart, K. (2006). Comprehensive study of tests for normality and symmetry: Extending the Spiegelhalter Test. *Journal of Statistical Computation and Simulation, 76*(9), 803–816.

Fatima, I., & Pathan, Z. H. (2016). Investigating learning strategies for vocabulary development: A comparative study of two universities of Quetta, Pakistan. *Advances in Language and Literary Studies, 7*(2). doi:10.7575/aiac.alls.v.7n.2p.7

Fatima, M., Hasan, K., Anwar, S., & Nawab, R. M. (2017). Multilingual author profiling on Facebook. *Information Processing and Management* (53), 886-904. Retrieved September 27, 2017, from www.elsevier.com/locate/infoproman

Fehlman, B. (2015). *Gender and dialogue differences in American novels.* Ouachita Baptist University. Retrieved February 22, 2018, from htp://scholarlycommons.obu.edu/ english_class_publications/6

Fehr, B. (1996). *Friendship processes.* Thousand Oaks, CA: Sage.

Filho, P. P., Pardo, T. A., & Alu´ısio, S. M. (2013). An evaluation of the Brazilian Portuguese LIWC dictionary for sentiment analysis. *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology* (pp. 215–219). Brazil: Sociedade Brasileira

de Computa¸c˜ao. Retrieved October 15, 2017, from http://www.aclweb.org/ anthology/ W13-4829

Firth, J. R. (1935). The technique of semantics. *Transactions of the Philological Society, 34*(1), 36-72. Retrieved October 18, 2017, from http://onlinelibrary.wiley.com/doi/10.1111/j. 1467-968X.1935.tb01254.x/pdf

Firth, J. R. (1968). *Selected papers of JR Firth: 1952-59.* Bloomington: Indiana University Press.

Fischer, A. (2011). Gendered social interactions in face-to-face and computer-mediated communication. In A. Kappas, & N. C. Kramer (Eds.), *Face-to-face communication over the internet: Emotions in a web of culture, language and technology* (pp. 53-77). Cambridge: Cambridge University Press.

Fischer, A. H. (1993). Sex differences in emotionality: Fact or stereotype? *Feminism and Psychology, 3*(3), 303-318. Retrieved January 10, 2018, from https://doi.org/10.1177/ 0959353593033002

Fishman, P. M. (1978). Interaction: The work women do. *Social Problems, 25*, 397–406.

Fishman, P. (1980). Conversational insecurity. In G. Howard, R. Robinson, & P. Smith (Eds.), *Language: Social psychological perspectives* (pp. 127-132). New York: Pergamon Press.

Fishman, P. M. (1983). Interaction: The work women do. In C. K. B. Thorne (Ed.), *Language, gender and society* (pp. 89-101). Boston: Heinle and Heinle Publishers.

Fishman, P. M. (1997). Interaction: The work women do. In N. Coupland, & A. Jaworski (Eds.), *Sociolinguistics: A reader* (pp. 416-429). New York: Macmillan Education.

Fivush, R., & Buckner, J. P. (2000). Gender, sadness, and depression: The development of emotional focus through gendered discourse. In A. Fischer (Ed.), *Gender and emotion: Social psychological perspectives* (pp. 232-253). Cambridge, UK.

Flekova, L., & Gurevych, I. (2013). Can we hide in the web? Large scale simultaneous age and gender author profiling in social media. *Notebook for PAN at CLEF 2013*. Retrieved March 18, 2018, from http://ceur-ws.org/Vol-1179/CLEF2013wn-PAN-FlekovaEt2013. pdf

Fletcher, W. (2007). Concordancing the web: Promise and problems, tools and techniques. In M. Hundt, C. Biewer, & N. Nesselhauf (Eds.), *Corpus linguistics and the web* (pp. 25-45). Amsterdam, Netherlands: Rodopi.

Flowerdew, L. (2004). The argument for using English specialized corpora to understand academic and professional language. In U. Connor, & T. A. Upton (Eds.), *Discourse in the professions* (pp. 11-36). Amsterdam, Netherlands: John Benjamins Publishing Co.

Flynn, F. J., Reagans, R. E., Amanatullah, E. T., & Ames, D. R. (2006). Helping one's way to the top: Self-monitors achieve status by helping others and knowing who helps whom. *Journal of Personality and Social Psychology, 91*(6), 1123-1137. doi:10.1037/0022-3514.91.6. 1123

Focault, M. (1972). *The archaeology of knowledge.* London: Tavistock.

Foucault, M. (1979). *Discipline and punish: The birth of the prison.* Harmondsworth: Penguin.

Foucault, M. (1981). The order of discourse. In Robert Young (Ed.) *Untying the text: A post-structuralist reader* (pp. 48-78). Boston: Routledge & Kegan Paul.

Francis, G. (1993). A corpus-driven approach to grammar: Principles, methods and examples. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology* (pp. 137–156). Amsterdam, Netherlands: John Benjamins. Retrieved October 19, 2017, from http:// ebookcentral.proquest.com

Friginal, E. (2009). *The language of outsourced call centers.* Amsterdam/Philadelphia: John Benjamin's Publishing.

Friginal, E., & Hardy, J. A. (2014). *Corpus-based sociolinguistics.* Oxon: Routledge.

Fuller, J. E. (2004). Equality in cyberdemocracy? Gauging gender gaps in online civic participation. *Social Science Quarterly, 85*(4), 938–957. Retrieved October 23, 2017, from http://onlinelibrary.wiley.com/doi/10.1111/j.0038-4941.2004.00252.x/epdf

Gao, F., & Martin-Kratzer, R. (2011, May 10). Gender differences in Chinese journalists' blogs. *Chinese Journal of Communication, 4*(2), 167–181. doi:10.1080/17544750.2011.565675

Galinsky, A. D., Magee, J. C., Gruenfeld, D. H., & Whitson, J. A. (2008). Power reduces the press of the situation: Implications for creativity, ocnformity, and dissonance. *Journal of Personality and Social Psychology, 95*(6), 1450-1466.

Galinsky, A. D., Magee, J. C., Inesi, M. E., & Gruenfeld, D. H. (2006). Power and perspectives not taken. *Psychological Science, 17*(12), 1068-1074.

Garden, M. (2011). Defining blog: A fool's errand or a necessary undertaking. *Journalism, 13*(4), 483–499. doi:10.1177/1464884911421700

Gatto, M. (2014). *The web as corpus.* London: Bloomsburry.

Gentry, J. W., & Doering, M. (1979). Sex role orientation and leisure. *Journal of Leisure Research, 11*, 102-111.

Gentry, M. (1989). Introduction feminist perspectives on gender and thought: Paradox and potential. In M. Crawford, & M. Gentry (Eds.), *Gender and thought: Psychological perspectives* (pp. 1-16). New York: Springer-Verlag.

Halliday, M. A. (1978). *Language as social semiotic: The social interpretation of language and meaning.* London: Edward Arnold.

Hamdan, S. M., & Hamdan, J. M. (2013). Authors' perceptions of author's gender: A myth or a truth? *International Journal of English and Literature, 4*(10), 523-528.

Hamdi, S., & Dabaghi, A. (2012). Gender Differences in Iranian EFL students' letter writing. *International Journal of Applied Linguistics and English Literature, 1*(7), 155-169.

Hamid, S., & Loewenthal, K. M. (1996). Inferring gender from handwriting in Urdu and English. *The Journal of Social Psychology, 136*(6), 778-782. doi:10.1080/00224545.1996.9712254

Hanafiyeh, M., & Afghari, A. (2014). Gender differences in the use of hedges, tag questions, intensifiers, empty adjectives, and adverbs: A comparative study in the speech of men and women. *Indian Journal of Fundamental and Applied Life Sciences, 4*(4), 1168-1177.

Hancock, A. B., & Rubin, B. A. (2015). Influence of communication partner's gender on language. *Journal of Language and Social Psychology, 1*, 46–64.

Harcourt, W. (2000). The personal and the political: Women using the Internet. *Cyberpsychology & Behavior, 3*(5), 693-697. doi:10.1089/10949310050191692

Harp, D., & Tremayne, M. (2006). The gendered blogosphere: Examining inequality using network and feminist theory. *Journalism & Mass Communication Quarterly, 83*(2), 247–264. Retrieved October 30, 2017, from https://search.proquest.com/docview/ 216931526? accountid=8630

Harrington, K. (2008). Perpetuating difference? Corpus linguistics and gendering of reported dialogue. In K. Harrington, L. L. Litosseliti, H. Sauntson, & J. Sunderland (Eds.), *Gender and language research methodologies* (pp. 85-102). Bassinstoke: Palgrave McMillan.

Harris, S. (1984). Questions as a mode of control in magistrates' courts. *International Journal of the Sociology of Language, 49*, 5-28.

Harrison, J. (1978). Warning: The male sex role may be dangerous to your health. *Journal of Social Issues*, 65-86.

Hartelius, E. J. (2005). A content-based taxonomy of blogs and the formation of a virtual community. *Kaleidoscope: A Graduate Journal of Qualitative Communication Research, 4*, 71-91.

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 *The sociology of the languages of American women* (pp. 81–90). San Antonio, Texas, USA: Trinity University Press.

Hawkins, D. T. (1996). Hunting, grazing, browsing: A model for online information retrieval. *Online, 20*(1), 71-73. Retrieved October 25, 2017, from https://search.proquest.com/ docview/199896999?accountid=8630

Heilman, M. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues, 57*, 657–674.

Herring, S. C. (1993). Gender and democracy in computer-mediated communication. *Electronic Journal of Communication*, 476-489. Retrieved from www.cios.org/getfile/ HERRING_ V3N293

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

Herring, S., Kouper, I., Scheidt, L., & Wright, E. (2004). *Women and children last: The discursive construction of weblogs.* Retrieved from Into the blogosphere: Rhetoric, community, and culture of weblogs.: http://blog.lib.umn.edu/blogosphere/women_and_children.html

Hewitt, H. (2005). *Blog: Understanding the information reformation that's changing your world.* Nashville: Thomas Nelson, Inc.

Heylighen, F., & Dewaele, J. M. (1999). *Formality of language: Definition, measurement and behavioral determinants.* Internal Report, Center "Leo Apostel", University of Brussels, Retrieved October 11, 2017, from http://pespmc1.vub.ac.be/Papers/Formality. pdf

Heylighen, F., & Dewaele, J. (2002). Variation in the contextuality of language: An empirical measure. *Foundations of Science, 7(3)*, 293-340. doi:https://doi.org/10.1023/ A:10196611 26744

Hiatt, M. (1977). *The way women write.* New York: Teachers College Press.

Hirschman, L. (1994). Female-male differences in conversational interaction. *Language in Society, 23*, 427-442.

Hirsh, J. B., & Peterson, J. B. (2009). Personality and language use in self-narratives. *Journal of Research in Personality, 43*, 524–527. doi:10.1016/j.jrp.2009.01.006

Holman, E., & Silver, R. (1998). Getting "stuck" in the past: Temporal orientation and coping trauma. *Journal of Personality and Social Psychology, 5*, 1146-1163.

Holmes, J. (1995). *Women, men and politeness.* Harlow: Longman.

Holmes, J. (2013). *An introduction to sociolinguistics* (4th ed.). Oxon, UK: Routledge.

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

Horváth, J. (2006). Keywords in context: Corpus Linguistics. In Borbála Richter (Ed.), *First steps in theoretical and applied linguistics* (pp. 111-118). Budapest, Hungary: Bölcsész Konzorcium.

Huaxue, W., & Dechang, J. (2014). Cross-gender differences on netspeak. *Canadian Social Science, 10*(3), 157-161.

Huber-Carol, C., Balakrishnan, N., Nikulin, M., & Mesbah, M. (2002). *Goodness-of-fit tests and model validity.* Boston: Birkhäuser.

Huffaker, D. A., & Calvert, S. L. (2005). Gender, identity, and language use in teenage blogs. *Journal of Computer-Mediated Communication, 10*(2), JCMC10211. https://doi.org/10.1111/j.1083-6101.2005.tb00238.x

Hunston, S. (2008). Collection strategies and design decisions. In A. Lüdeling, & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (pp. 154-167). Berlin, Germany: Walter de Gruyter.

Hussain, M. N., & Afsar, Dr. A., (2010). Language and gender—linguistic analysis of intermediate English textbooks in Pakistan. *Language in India, 10*, 26–42.

Hussain, F., & Rasul, S. (2016). Sarcasm, humour and exaggeration in Pakistani Punjabi wedding songs: Implicit gendered identities. *Pakistan Journal of Women's Studies: Alam-e-Niswan, 23*(2), 21-44. Retrieved November 5, 2017, from https://search.proquest.com/openview/911480da64c1b9ffc825e7eca02be6a8/1?pq-origsite=gscholar&cbl=40072

Iachini, T., Sergi, I., Ruggiero, G., & Gnisci, A. (2005). Gender differences in object location memory in a real three-dimensional environment. *Brain and Cognition, 59*, 52-59.

Ikram, M., & Ghani, M. (2013). Gender and language learning motivation. *Academic Research International, 4*(2), 536–540.

Iosub, D., Laniado, D., Castillo, C., Fuster Morell, M., & Kaltenbrunner, A. (2014). Emotions under discussion: Gender, status and communication in online collaboration. *PloS One, 9*(8).

Isaac, C., Chertof, J., Lee, B., & Carnes, M. (2011). Do students' and authors' genders affect evaluations? A linguistic analysis of medical student performance evaluations. *Academic Medicine, 86*(1), 59-66.

Jabeen, S., Chaudhary, A. Q., & Omar, S. (2014). Gender discrimination in curriculum: A reflection from Punjab textbook board. *Bulletin of Education and Research, 3*(1), 55–57.

Jackson, L. A., Ervin, K. S., Gardner, P. D., & Schmitt, N. (2001). Gender and the internet: Women communicating and men searching. *Sex Roles, 5*(6), 363–378. doi:10.1023/A: 1010937901821

Jaffe, M., Lee, Y., Huang, L., & Oshagan, H. (1999). Gender, pseudonyms, and CMC: Masking identities and baring souls. *Gender, Language and CMC, 15*, 221-234.

James, D., & Clarke, S. (1993). Women, men, and interruptions: A critical review. In D. Tannen (Ed.), *Gender and conversational interaction* (pp. 231-280). Oxford: Oxford University Press.

Janmohamed, S. Z. (2010, June). Rebuilding Islam's brand. *Emel* (69), n.p. Retrieved July 14, 2018, from http://www.emel.com/article?id=73&a_id=2047

Javed, M., Juan, W. X., & Nazli, S. (2013). A study of students' assessment in writing skills of the English language. *International Journal of Instruction, 6*(2), 129-144. Retrieved November 4, 2017, from https://eric.ed.gov/?id=ED544075

Jespersen, O. (1922). *Language: Its nature, development and origin* (First ed.). London, UK: George Allen & Unwin Ltd. Retrieved October 02, 2017, from https://archive.org/details/languageitsnatur00jespiala

Jie, H. W. (2015). *Corpus-driven analysis of male and female weblog users in an ESL context* (Doctoral Dissertation), University of Malaya, Faculty of Languages and Linguistics, Kuala Lumpur.

Jieun, J., & Jae-Woong, C. (2009). A keyword analysis of English intensifying adverbs in male and female speech in ICE-GB. *Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation*, (pp. 210-219).

Jingwei, L. (1998). Language gender-related differences and its cause analysis. *Shandong Foreign Language Teaching*, 12-16.

Jinlan, H., Chung, C. K., Hui, N. H., Yizheng, L., Yitai, X., Lam, B. C., & Weiquan, C. (2012). The development of the Chinese Linguistic Inquiry and Word Count Dictionary. *Chinese Journal of Psychology, 54*(2), 185-201. Retrieved October 14, 2017, from https://www.researchgate.net/profile/Yi_Cheng_Lin4/publication/303135549_Development_of_the_Chinese_linguistic_inquiry_and_word_count_dictionary/links/58ca9505a6fdcc1d1fea98c9/Development-of-the-Chinese-linguistic-inquiry-and-word-count-dictionary. pdf

Jinyu, D. (2014). Study on gender differences in language under the sociolinguistics. *Canadian Social Science, 10*(3), 92-96.

Jeon, J., & Choe, J. (2009). A Key Word Analysis of English Intensifying Adverbs in Male and Female Speech. *ICE-GB. 23rd Pacific Asia Conference on Language, Information and Computation*, (pp. 210–219).

Johnsen, J.-A. K., Vambheim, S. M., Wangberg, R. W., & C, S. (2014). Language of motivation and emotion in an Internet support group for smoking cessation: Explorative use of

automated content analysis to measure regulatory focus. *Psychology Research and Behavior Management, 7*, 19–29. Retrieved October 15, 2017, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3896322/pdf/prbm-7-019.pdf

Johnson, F. L., & Aries, E. J. (1983). Conversational patterns among same-sex pairs of late-adolescent close friends. *The Journal of Genetic Psychology, 142*(2), 225-238. doi:http://dx.doi.org/10.1080/00221325.1983.10533514

Johnson, J. T., & Shulman, G. A. (1988). More alike than meets the eye: Perceived gender differences in subjective experience and its display. *Sex Roles, 1*(2), 67-79. Retrieved January 10, 2018, from https://doi.org/10.1007/BF00292465

Johnson, S., Culpeper, J., & Suhr, S. (2003). From "politically correct councillors" to "Blairite nonsense": Discourses of political correctness in three British newspapers. *Discourse and Society, 14*(1), 28–47. Retrieved October1 17, 2017, from http://journals.sagepub.com/doi/pdf/10.1177/0957926503014001928

Johnson, S. (1997). Theorizing language and masculinity: A feminist perspective. In S. Johnson, & U. H. Meinhof (Eds.), *Language and masculinity* (pp. 8-26). Cambridge, Massachesetts, USA: Blackwell Publishers.

Johnson, Z. (2017). *How many websites and blogs are there on the internet?* Retrieved October 26, 2017, from www.blogging.org: https://blogging.org/blog/how-many-websites-and-blogs-are-on-the-internet/

Jones, J. J. (2016). Talk "like a man": The linguistic styles of Hillary Clinton, 1992–2013. *American Political Science Association, 14*(3), 625-641. Retrieved January 18, 2018, from https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1537592716001092

Jordan, J., Surrey, J., & Kaplan, A. (1991). Women and empathy: Implications for psychological development and psychotherapy. In J. Jordan, A. Kaplan, M. J.B., I. Stiver, & J. Surrey (Eds.), *Women's growth in connection: Writings from the Stone Center* (pp. 27-50). New York: Guilford Press.

Jule, A. (2005). Introduction: The meeting of gender, language and religion. In A. Jule (Ed.), *Gender and the language of religion* (pp. 1-6). New York: Palgrave Macmillan.

Junghaenel, D., Smyth, J. M., & Santner, L. (2008). Linguistic dimensions of psychopathology. *Journal of Social & Clinical Psychology, 27*, 36-55.

Kabat-Zinn, J. (1982). An outpatient program in behavioural medicine for chronic pain patients based on the practice of mindfulness meditation: Theoretical considerations and preliminary results. *General Hospital Psychiatry, 1*, 33-47.

Kabat-Zinn, J. (1994). *Wherever you go. There you are: Mindfulness meditation in everyday life.* New York: Hachette Books.

Kalman, Y. M., & Gergle, D. R. (2010). CMC cues enrich lean online communication: The case of letter and punctuation mark repetitions. *Proceedings of the Fifth Mediterranean Conference on Information Systems.* Elsevier.

Kamran, S., & Mansoor, S. (2017). Globalization and language use on social media in Pakistan. *European Journal of Language and Literature Studies, 8*(1). doi: http://dx.doi.org/ 10.26417/ejls.v8i1.p79-84

Kapidžić, S., & Herring, S. (2011). Gender, communication, and self-presentation in teen chatrooms revisited: Have patterns changed? *Journal of Computer-Mediated Communication, 17*(1), 39-59.

Karlsson, S. (2007). *Gender-related differences in language use* (Unpublished Master's thesis). Luleå University of Technology, Luleå, Sweden.

Kaukab, S. R., & Mehrunnisa, S. (2014). Blogging: An effective medium for communicating and discussing latest news amongst youth of Karachi. *Academic Research International, 5*(4). Retrieved December 11, 2017, from http://www.savap.org.pk/journals/ARInt./Vol.5(4)/ 2014(5.4-35).pdf

Kelly, J. i., & -Comeaux, S. L. (1999). Gender-emotion stereotypes are context specific. *Sex Roles, 40*(112), 107-120. Retrieved January 10, 2018, from https://doi.org/10.1023/ A:101883450 1996

Kemppanen, H. (2004). Keywords and ideology in translated history texts: A corpus-based analysis. *Across Languages and Cultures, 5*(1), 89–106. doi:10.1556/Acr.5.2004.1.5

Kendal, S. (2017). *STEEMIT: Total number of blogs*. Retrieved October 26, 2017, from www.steemit.com: https://steemit.com/steemit/@stephenkendal/steemit-total-number-of-blogs-published-in-2017-currently-stands-at-144-757-average-number-of-blogs-per-day-is-1-392

Kendall, S., & Tannen, D. (2001). Discourse and gender. In D. Schiffrin, D. Tannen, and H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 548–567). Oxford: Blackwell Publishers.

Keng, S. L., Smoski, M. J., & & Robins, C. J. (2011). Effects of mindfulness on psychological health: A review of rmpirical studies. *Clinical Psychology Review, 31*, 1041-1056. doi:http://dx.doi.org/10.1016/j.cpr.2011.04.006

Keskin, S. (2006). Comparison of several univariate normality tests regarding Type I Error rate and power of the test in simulation based small samples. *Journal of Applied Science Research, 2*(5), 296-300.

Key, M. R. (1975). *Male/female language.* Metuohen, N.J.: The Scarecrow Press, Inc.

Khan, H. I. (2012). The evolution of Pakistani English (PakE) as a legitimate variety of English. *International Journal of Applied Linguistics & English Literature, 1*(5), 90-99. Retrieved October 9, 2017, from http://www.journals.aiac.org.au/index.php/IJALEL/article/view/ 74 7/677

Khan, Q., Sultana, N., Bughio, Q., & Naz, A. (2014). Role of language in gender identity formation in Pakistani school textbooks. *Indian Journal of Gender Studies, 21*(1), 55–84. doi:10.1177/0971521513511200

Kiefer, I., Rathmanner, T., & Kunze, M. (2005). Eating and dieting differences in men and women. *The Journal of Men's Health & Gender, 2*(2), 194–201.

Kilgarriff, A., & Grefenstette, G. (2003). Introduction to the special issue on the web as corpus. *Computational Linguistics, 29*(3), 333-347. doi:10.1162/089120103322711569

Kimura, D. (1992, September). Sex differences in the brain. *Scientific American, 267*, 119-125.

Kirk, U. (1992). Evidence for early acquisition of visual organization ability: A developmental study. *The Clinical Neuropsychologist, 6*, 171-177.

Kirkpatrick, L. A., & Epstein, S. (1992). Cognitive-experiential self-theory and subjective probability: Further evidence for two conceptual systems. *Journal of Personality and Social Psychology, 63*(4), 534-544. Retrieved January 8, 2018, from http://citeseerx.ist. psu.edu/viewdoc/download?doi=10.1.1.473.3747&rep=rep1&type=pdf

Kivran-Swaine, F., Brody, S., Diakopoulos, N., & Naaman, M. (2012). Of joy and gender: Emotional expression in online social networks. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion*, (pp. 139–142). Seatle, Washington, USA.

Kjellmer, Goeran (1986) 'The lesser man': Observations on the role of women in modern English writings. In Jan Arts and Willem Meijs (eds) *Corpus Linguistics II* (pp. 163–176). Amsterdam: Rodopi.

Klest, B., & Freyd, J. J. (2007). Global ratings of essays about trauma: Development of the GREAT code, and correlations with physical and mental health outcomes. *Journal of Psychological Trauma, 6*(1), 1-20. doi:1932-2895

Kolsky, S. D. (2012). Making examples of women: Juan Luis Vives' the education of a Christian woman. *Early Modern Culture Online, 3*(1), 14-38. Retrieved October 6, 2017, from journal.uia.no/index.php/EMCO/article/download/41/33

Koppel, M., Argamon, S., & Shimoni, A. R. (2002). Automatically categorizing written text by author gender. *Literary and Linguistic Computing, 17*(4), 401-412. Retrieved October 12, 2017, from https://watermark.silverchair.com/170401.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAe0wggHpBgkqhkiG9w0BBwagggHaIIB1gIBADCCAc8GCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMOm3rzQ46OhWoPy99AgEQgIIBoEnDUte_6dRUfOoldSwwkaUSOBqSMoCHoI14QBKGoSqrffvB

Koteyko, N. (2014). *Language and politics in post-Soviet Russia: A Corpus-assisted approach.* New York: Palgrave Macmillan.

Kousar, S. (2013). Oh, I am sorry': The realization patterns of apologizing in Urdu. *Journal of Research in Social Sciences, 2*(1), 70-82. Retrieved November 3, 2017, from https://search.proquest.com/docview/1536047435/fulltextPDF/AFF01710417E441FPQ/1?accountid=8630

Kramarae, C., & Taylor, H. J. (1993). Women and men on electronic networks: A conversation or a monologue? In C. K. H. J. Taylor, & M. Ebben (Eds.), *Women, information technology, and scholarship* (pp. 52–61). Urbana, IL: Center for Advanced Study.

Kramer, C. (1974). Folk-linguistics: Wishy-washy mommy talk. *Psychology Today*, 82-89.

Kring, A. M., & Gordon, A. H. (1998). Sex differences in emotion: Expression, experience, and physiology. *Journal of Personality and Social Psychology, 74*(3), 686-703. Retrieved January 10, 2018, from http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=C1F169C2AE49009E990CE3A0570F2291?doi=10.1.1.379.5826&rep=rep1&type=pdf

Ku, P., & Yang, A. (1999). An analysis of key words in tourism English. *The Proceedings of the 1999 English for Specific Purposes Conference* (pp. 232–241). Taoyuan, Taiwan: Ming Chuan University.

Kučera, J., & Francis, W. (1967). *Computational analysis of present-day American English.* Providence: Brown University Press.

Kuo, S. (2003). Involvement vs detachment: Gender differences in the use of personal pronouns in televised sports in Taiwan. *Discourse Studies, 5*(4), 479-494. Retrieved January 27, 2018, from https://doi.org/10.1177/14614456030054002

Labov, W. (1966). *The social stratification of English in New York city.* Washington: CAL.

Labov, W. (1972a). *Sociolinguistic patterns.* Philadelphia, PA: University of Pennsylvania Press.

Labov, W. (1972b). The logic of nonstandard English. In P. Giglioli (Ed.), *Language and social context* (pp. 179-215). Harmondsworth: Penguin.

Labov, W. (1990). "The intersection of sex and social class in the course of linguistic change". *Language Variation and Change 2*: 205-254.

Labov, W. (2001). *Principles of linguistic change: Social factors* (Vol. 2). Malden, MA: Blackwell.

ladette. (2017). *Dictionary*. (Oxford) Retrieved October 12, 2017, from English Oxford Living Dictionaries: https://en.oxforddictionaries.com/definition/ladette

Lahiri, S. P., & Lu, X. (2011). Informality judgment at sentence level and experiments with formality score. In G. A. (Ed.), *Computational linguistics and intelligent text processing: Lecture notes in Computer Science 6609* (pp. 446–457). Heidelberg, Berlin, Germany: Springer-Verlag.

Lakoff, R. (1973). Language and women's place. *Language in Society, 2*(1), 45-80. Retrieved October 02, 2017, from http://www.jstor.org/stable/4166707

Lakoff, R. T. (1975). *Language and women's place.* New York: Harper & Row.

Lakoff, R. T. (2004). Author's introduction: Language and women's place revisited. In M. Bucholtz (Ed.), *Language and women's place: Text and commentaries* (pp. 15-28). New York: Oxford University Press.

Lapierre, S., Bouffard, L., & Bastin, E. (1997). Personal goals and subjective well-being in later life. *International Journal of Aging and Human Development*, 287-303.

Larose, D. T. (2015). *Discovering statistics.* London, UK: W. H. Freeman & Co Ltd.

Laviosa, S., Pagano, A., Kemppanen, H., & Ji, M. (2017). *Textual and contextual analysis in empirical translation studies.* Singapore City, Singapore: Springer.

Leaper, C., & Ayres, M. (2007). A meta-analytic review of gender variations in adults' language use: Talkativeness, affiliative speech, and assertive speech. *Personality and Social Psychology Review, 11*(4), 328–63.

Leaper, C., & Robnett, R. (2011). Women are more likely than men to use tentative language, aren't they? A meta-analysis testing for gender differences and moderators. *Psychology of Women Quarterly, 35*(1), 129-142.

Leaper, C., & Smith, T. E. (2004). A meta-analytic review of gender variations in children's language use: Talkativeness, affiliative speech, and assertive speech. *Developmental Psychology*, 40, 993–1027.

Lee, C. (2003). How does instant messaging affect interaction between the genders. Stanford, CA: *The Mercury Project for Instant Messaging Studies at Stanford University*. Retrieved August, 11, 2017.

Lee, C. H., Park, J., & Seo, Y. S. (2006). An analysis of linguistic styles by inferred age in TV dramas. *Psychological Reports, 99*(2), 351-356. doi:10.2466/pr0.99.2.351-356

Leech, G. (1991). The state of the art in corpus linguistics. In K. Aijmer, & B. Altenberg (Eds.), *English corpus linguistics: Studies in honour of Jan Svartvik* (pp. 8-29). London, UK: Longman.

Lenard, D. B. (2016). *Gender differences in the political speeches from the 113th United States Congress.* University of Osijek, Faculty of Humanities and Social Sciences. Osijek: Unpublished PhD Dissertation. Retrieved from www.ffos.unios.hr/download/disertacija.pdf

Lenard, D. B. (2017). Gender differences in the personal pronouns usage on the corpus of congressional speeches. *Journal of Research Design and Statistics in Linguistics and Communication Science, 3*(2), 161–188. Retrieved February 01, 2018, from https://bib.irb.hr/datoteka/861053.Gender_differences_in_the_personal_pronouns_usage_of_the_corpus_of_congressional_speeches.pdf

Lester, D. (2004). *Katie's diary: Unlocking the mystery of a suicide.* London: Routledge.

Lester, D. (2014). *The "I" of the storm: Understanding the suicidal mind.* Berlin: De Gruyter Open Ltd.

Lester, D., & Leenaarsb, A. (2016). A comparison of suicide notes written by men and women. *Death Studies, 40*(3), 201–203. doi:0.1080/07481187.2015.1086449

Levy, M. (2009). WEB 2.0 implications on knowledge management. *Journal of Knowledge Management, 13*(1), 120-134.

Li, C.-s. R., Huang, C.-Y., Lin, W.-y., & Sun, C.-W. V. (2007). Gender differences in punishment and reward sensitivity in a sample of Taiwanese college students. *Personality and Individual Differences, 43*, 475–483.

Li, G., & Papadopoulos, A. (2002). A note on goodness of fit test using moments. *Statistica, 62*(1), 72–86.

Li, J. (2014). A sociolinguistic study of language and gender in desperate housewives. *Theory and Practice in Language Studies*, 52-57.

Lin, C.-W., Lin, M.-J., Wen, C.-C., & Chu, S.-Y. (2016). A word-count approach to analyze linguistic patterns in the reflective writings of medical students. *Medical Education Online, 21*, 1-7. doi:10.3402/meo.v21.29522

Lisheng, X. (1997). Studies on gender-related differences in discourse feature. *Journal of Foreign Languages*, 03-06.

Liu, L., Tang, L., Dong, W., Yao, S., & Zhou, W. (2016). An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus, 5*(1), 1608-1630. doi:10.1186/ s40064-016-3252-8

Llamas, C., Mullany, L., & Stockwell, P. (2007). Introduction. In C. Llamas, L. Mullany, & P. Stockwell (Eds.), *The Routledge companion to sociolinguistics* (pp. xv-xviii). Oxford: Routledge.

Ludu, P. S. (2014). Inferring gender of a Twitter user using celebrities it follows. Retrieved from https://www.researchgate.net/publication/262640860_Inferring_gender_of_a_Twitter_us er_using_celebrities_it_follows

Lutz, C. (1990). Engendered emotion: Gender, power, and the rhetoric of emotional control in American discourse. In C. Lutz, & L. Abu-Loghod (Eds.), *Language and the politics of emotion.*

MacKillop, J., & Anderson, E. J. (2007). Further psychometric validation of the Mindful Attention Awareness Scale (MAAS). *Journal of Psychopathology and Behavioral Assessment, 29*(4), 289-293.

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences.* Stanford, CA: Stanford University Press.

Maiwald, P. (2011). Exploring a corpus of George MacDonald's fiction. *North Wind: Journal of George MacDonald Studies, 30*, 50-84. Retrieved October 23, 2017, from http://www.snc. edu/northwind/documents/By_genre_or_topic/Language/Exploring_a_Corpus_of_Georg e_MacDonald's_Fiction_-_Patrick_Maiwald.pdf

Malik, I. H., & Akhtar, M. (2015). Differences in the use of English vocabulary learning strategies at higher secondary level. *Journal of Educational Research, 18*(1), 96-105. Retrieved November 2, 2017, from https://search.proquest.com/docview/1794151046/ fulltextPDF/ 2BB02EB4407A4CFBPQ/1?accountid=8630

Malone, M. (1997). *Worlds of talk: The presentation of self in everyday conversation.* Oxford: Polity Press.

Maltz, D. N., & Broker, R. A. (1982). A cultural approach to male-female miscommunication. In J. J. Gumperz (Ed.), *Language and Social Identity* (First ed., pp. 195-216). Cambridge, UK: Cambridge University Press.

Manjavacas, E. (2015). Statistical description of gender-related differences in language use in a Dutch blog corpus. Retrieved November 8, 2017, from https://www.researchgate.net/ publication/273462395/download

Mansoor, Z. (2013). Print media language: Contributing to the stereotypical portrayal of Pakistani women. *American International Journal of Contemporary Research*, 148-156. Retrieved November 5, 2017, from http://www.aijcrnet.com/journals/Vol_3_No_7_July_ 2013/18. pdf

Marmolejo-Ramos, F., & González-Burgos, J. (2013). A power comparison of various tests of univariate normality on ex-gaussian distributions. *Methodology: European Journal of Behavioral and Social Sciences, 9*(4), 137-149. Retrieved January 6, 2018, from https://doi. org/ 10.1027/1614-2241/a000059

Martin Rojo, L., & Gomez Esteban, C. (2005). The gender of power: The female style in labour organization. In M. Lazar (Ed.), *Feminist critical discourse analysis* (pp. 61-89). New York: Palgrave Macmillan.

Mass'o, G., Lambert, P. P., Rodr'ıguez, C., & Saur'ı, R. (2013). Generating new LIWC dictionaries by triangulation. In R. E. Banchs, F. Silvestri, T.-Y. Liu, M. Zhang, S. Gao, & J. Lang (Ed.), *Information retrieval technology: 9th Asia Information Retrieval Societies Conference AIRS 2013 Singapore, December 9-11, 2013. Proceedings* (pp. 263-271). London: Springer.

McAdams, M. (1995). Inventing an online newspaper. *Interpersonal Computing and Technology: An Electronic Journal for the 21st Century, 3*(3), 64-90.

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

McConnell-Ginet, S. (2011). *Gender, sexuality, and meaning: Linguistics practice and politics.* New York: Oxford University Press.

McEnery, T. (2006). *Swearing in English.* London: Routledge.

McEnery, T., & Wilson, A. (1996). *Corpus linguistics.* Edinburgh: Edinburgh University Press.

McEnery, T., Xiao, R., Tono, Y. (2006). *Corpus-based language studies*: *An advanced resource book.* Routledge. Retrieved November 06, 2019, from https://www.lancaster.ac.uk/fass/ projects/corpus/ZJU/xCBLS/CBLS.htm

McFadyen, R. G. (1996). Gender, status and 'powerless' speech: Interactions of students and lecturers. *British Journal of Social Psychology, 35*, 353-367.

McMillan, J., Clifton, A., McGrath, D., & Gale, W. (1977). Women's language: Uncertainty or interpersonal sensitivity and emotionality. *Sex Roles, 3*(6), 545-559.

Mehl, M. R., & Gill, A. J. (2010). Automatic text analysis. In S. D. Gosling (Ed.), *Advanced methods for conducting online behavioural research* (pp. 109-127). Washington, DC: American Psychological Association.

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology, 90*(5), 862–877.

Mehl, M., & Pennebaker, J. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology, 84*, 857–870.

Mehler, A., & Gleim, R. (2005). Polymorphism in generic web units. A Corpus linguistic study. Retrieved October 25, 2017, from https://pdfs.semanticscholar. org/0ce1/ 091fe2a7ff54e 80fd5055bf7cf59842f7c98.pdf

Mehl, M. R. (2006). Quantitative text analysis. In M. Eid, & D. Ed (Eds.), *Handbook of multimethod measurement in psychology* (pp. 141-156). American Psychological Association (APA).

Mendes, M., & Pala, A. (2003). Type I Error rate and power of three normality tests. *Pakistan Journal of Information and Technology, 2*(2), 135-139.

Metlo, Z. (2012). *The perception of honour among the British-Pakistani community in Watford, United Kingdom* (Doctoral dissertation), School of Modern Languages and Cultures, University of Leeds.

Meyerhoff, M. (2006). *Introducing sociolinguistics.* Oxford, UK: Routledge.

Werlen, L. M. (2015). Statistical learning methods for profiling analysis. In *Conference handbook of empirical methods in natural language processing* (n.p.). Lisbon: Association for Computational Linguistic.

Mikros, G. (2013). Systematic stylometric differences in men and women authors: A corpus-based study. *Issues in Quantitative Linguistics, 3*, 206-223.

Mills, S. (2003). *Gender and politeness.* Cambridge, UK: Cambridge University Press.

Mills, S. (2004). *Discourse* (2nd ed.). London: Routledge.

Milroy, J., & Milroy, L. (1978). Belfast: Change and variation in an urban vernacular. In P. Trudgill (Ed.), *Sociolinguistic patterns in British English* (pp. 19–36). London: Arnold.

Milroy, L. (1980). *Language and social networks.* Oxford: Blackwell.

Missagia, S. V., Oliveira, S. R., & Rezende, D. C. (2013). Beauty and the beast: Gender differences in food-related behavior. *Revista Brasileira de Marketing - REMark, São Paulo, 12*(1), 149-165.

Mita, A. (2004). Voices of the marginalized on the internet: Examples from a website for women of South Asia. *Journal of Communication, 54*(3), 492-510. doi:10.1111/j.1460-2466.2004.tb02641.x

Mondorf, B. (2002). Gender differences in English syntax. *Journal of English Linguistics, 30*(2), 158-180.

Moore, H. T. (1922). Further data concerning sex differences. *Journal of Abnormal Psychology, 17*, 210-214.

Mukherjee, A., & Liu, B. (2010). Improving gender classification of blog authors. *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing 9-11 October 2010* (pp. 207–217). MIT, Massachusetts, USA: Association for Computational Linguistics.

Mulac, A. (1989). Men's and women's talk in same-gender and mixed-gender dyads: Power or polemic? *Journal of Language and Social Psychology, 8*(3-4), 249-270. Retrieved January 11, 2018, from https://doi.org/10.1177/0261927X8983006

Mulac, A., & Bradac, J. (1995). Women's style in problem-solving interactions: Powerless or simply feminine? In *Gender, Power, and Communication in Human Relationships* (pp. 83-104). Hillsdale, NJ: Erlbaum.

Mulac, A., & Lundell, T. (1994). Effects of gender-linked language differences in adults' written discourse: Multivariate tests of language effects. *Language and Communication, 14*(3), 299-309. Retrieved January 11, 2018, from https://doi.org/10.1016/0271-5309(94)90007-8

Mulac, A., & Lundell, T. L. (1980). Differences in perceptions created by syntactic-semantic productions of male and female speakers. *Communication Monographs, 47*(2), 111-118. doi:10.1080/03637758009376024

Mulac, A., & Lundell, T. L. (1986). Linguistic contributors to the gender-linked language effect. *Journal of Language & Social Psychology, 5*, 81–101.

Mulac, A., Giles, H., Bradac, J., & Palomares, N. (2013). The gender-linked language effect: An empirical test of a general process model. *Language Sciences, 38*, 22–31. Retrieved January 11, 2018, from https://doi.org/10.1016/j.langsci.2012.12.004

Mulac, A., Lundell, T., & Bradac, J. (1986). Male/female language differences and attributional consequences in a public speaking situation: Toward an explanation of the gender-linked language effect. *Communication Monographs, 53*, 115-129.

Mulac, A., Seibold, D., & Farris, J. (2000). Female and male managers' and professionals' criticism giving: Differences in language use and effects. *Journal of Language and Social Psychology, 19*(4), 389–415. Retrieved January 11, 2018, from https://doi.org/10.1177/0261927X00019004001

Mulac, A., Studley, L., & Blau, S. (1990). The gender-linked language effect in primary and secondary students' impromptu essays. *Sex Roles, 23*(9), 439-470. Retrieved January 11, 2018, from https://search-proquest-com.ezpr oxyd.bham.ac.uk/docview/1308100138?accountid=8630

Mulac, A., Wiemann, J., Widenmann, S., & Gibson, T. (1988). Male/female language differences and effects in same-sex and mixed-sex dyads: The gender-linked language effect. *Communication Monographs, 55*(4), 315-335.

Myerhoff, B. (1980). *Number our days.* New York: Simon & Schuster.

Nagarajan, M., & Hearst, M. (2009). An examination of language use in online dating profiles. *Proceedings of the Third International ICWSM Conference on Weblogs and Social Media May 17-20, 2009* (pp. 266-269). Association for the Advancement of Artificial Intelligence. San Jose, California, USA: AAAI Press.

Nathanson, C. A. (1977). Sex roles as variables in preventive health behavior. *Journal of Community Health, 3*(2), 142–155.

Naveed, A. (2014). Analysis of cyber language: Identifying gender boundaries. *European Academic Research. 2*(1), 9706–9724.

Nazir, B. (2012). Gender patterns on Facebook: A sociolinguistic perspective. *International Journal of Linguistics, 4*(3), 252-265. Retrieved November 3, 2017, from http://www.macrothink.org/ journal/index.php/ijl/article/view/1899/pdf

Ndambuki, J., & Janks, H. (2010). Political discourses, women's voices: Mismatches in representation. *Critical Approaches to Discourse Analysis across Disciplines, 4*(1), 73-92. Retrieved January 24, 2018, from http://www.lancaster.ac.uk/fass/journals/cadaad/wp-content/uploads/2015/01/Volume-4_Ndambuki__Janks.pdf

Neideen, T., & Brasel, K. (2007). Understanding statistical tests. *Journal of Surgical Education, 64*(2), 93-96. Retrieved January 6, 2018, from http://osp.mans.edu.eg/ tmahdy/ papers_ of_ month/0706_statistical.pdf

Netspeak. (2003). *Microsoft internet & networking dictionary.* Redmond, Washington: Microsoft Press.

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. *Discourse Processes, 45*(3), 211-236. Retrieved January 2, 2018. doi: 10.1080/01638530802073712

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin, 29*(5), 665-675. doi:10.1177/0146167203251529

Noreen, A., & Zubair, S. (2012). Gender and time-out-talk: An analysis of construction of femininities by young Pakistani women. *Pakistan Journal of Women's Studies: Alam-e-Niswan, 19*(1), 95-114. Retrieved November 3, 2017, from https://search.proquest.com/ docview/1138813562?accountid=8630&rfr_id=info%3Axri%2Fsid%3Aprimo

Noughabi, H., & Arghami, N. (2011). Monte Carlo comparison of seven normality tests. *Journal of Statistical Computation and Simulation, 81*(8), 965–972.

Nowson, S. (2006). *The language of weblogs: A study of genre and individual differences* (Doctoral dissertation), Institute for Communicating and Collaborative Systems, University of Edinburgh, School of Informatics.

O' Barr, W., & Atkins, B. (1998). Women's language" or "powerless language? In J. Coates (Ed.), *Language and gender: A reader.* Oxford, UK: Blackwell Publishing.

O'Barr, W. M. (1982). *Linguistic evidence: Language, power, and strategy in the courtroom.* New York: Academic Press.

O'Brien, R., Hunt, K., & Hart, G. (2005). 'It's caveman stuff, but that is to a certain extent how guys still operate': Men's accounts of masculinity and help seeking. *Social Science and Medicine, 61*, 503–516.

O'Kearney, R., & Dadds, M. (2004). Developmental and gender differences in the language for emotions across the adolescence years. *Cognition and Emotion, 18*(7), 913-938.

O'Reilly, T. (2007). What is Web 2.0: Design patterns and business models for the next generation of software. *Communication & Strategies*, 17-37. Retrieved October 22, 2017, from https:// mpra.ub.uni-muenchen.de/4578/1/MPRA_paper_4578.pdf

Oberlander, J., & Gill, A. J. (2006). Language with character: A stratified corpus comparison of individual differences in e-mail communication. *Discourse Processes, 42*(3), 239–270. Retrieved October 28, 2017, from http://dx.doi.org/10.1207/s15326950dp4203_1

Ochs, E. (1992). Indexing gender. In A. Duranti and C. Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon* (pp. 335-358). Cambridge: Cambridge University Press.

Ogan, C., Çiçek, F., & Özakça, M. (2005). Letters to Sarah: Analysis of email responses to an online editorial. *New Media & Society, 7*(4), 533-557. doi:10.1177/1461444805054117

Opina, K. G. (2017). Verbal communication behaviors: How male and female university students interact in gendered talks. *International Journal of Language and Linguistics, 5*(5), 135-142.

Owen, P. R., & Padron, M. (2015). The language of toys: Gendered language in toy advertisements. *Journal of Research on Women and Gender, 6*, 67-80. Retrieved October 13, 2017, from https://journals.tdl.org/jrwg/index.php/jrwg/article/view/24

Page, R. (2011). Blogging on the body: Gender and narrative. In R. Page, & B. Thomas (Eds.), *New narratives: Stories and storytelling in the digital age* (pp. 220-238). London, UK: University of Nebraska Press.

Palander-Collin, M. (1998). Grammaticalization of "I think" and "methinks" in Late Middle and Early Modern English: A sociolinguistic perspective. *Neuphilologische Mitteilungen, 99*(4), 419-442. Retrieved January 24, 2018, from http://www.jstor.org/stable/43345083

Palander-Collin, M. (1999). I Think, Methinks: Register variation, stratification, education and nonstandard language. In I. Taavistsainen, G. Melchers, & P. Patha (Eds.), *Writing in nonstandard English* (pp. 243-262). Amsterdem: John Benjamin Publishing Company.

Palomares, N. A. (2008). Explaining gender-based language use: Effects of gender identity salience on references to emotion and tentative language in intra- and intergroup contexts. *Human Communication Research, 34*(2), 263-286.

Palomares, N. A. (2009). Women are sort of more tentative than men, aren't they?: How men and women use tentative language differently, similarly, and counterstereotypically as a function of gender salience. *Communication Research, 4*, 538-560.

Parkes, M. B. (2016). *Pause and effect: An introduction to the history of punctuation in the West.* Oxon: Routledge.

Parkins, R. (2012). Gender and emotional expressiveness: An analysis of prosodic features in emotional expression. *Griffith Working Papers in Pragmatics and Intercultural Communication, 5*(1), 46-54.

Parsons, T., Larson, P., Kratz, K., Thiebaux, M., Bluestein, B., Buckwalter, J., & Rizzo, A. (2004). Sex differences in mental rotation and spatial rotation in a virtual environment. *Neuropsychologia, 42*, 555-562.

Pearce, M. (2008). Investigating the collocational behaviour of MAN and WOMAN in the BNC using Sketch Engine. *Corpora, 3*(1), 1–29.

Peersman, C., Daelemans, W., & Vaerenbergh, L. V. (2011). Predicting age and gender in online social networks. *3rd International Workshop on Search and Mining User-generated Contents (SMUC 2011).* Glasgow: ACM Digital Library. Retrieved January 11, 2018, from http://ir.ii.uam.es/smuc2011/submissions.html

Pennebaker, J. (2013). *The secret life of pronouns: What our words say about us.* New York: Bloomsbury Press.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77*(6), 1296-1312. doi:10.1037/ 0022-3514.77.6.1296

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PloS One, 9*(12), 1-10. Retrieved January 8, 2018, from http://journals.plos.org/plosone/ article/file?id= 10.1371/journal.pone.0115844&type=printable

Pennebaker, J., Boyd, R., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015.* Austin, TX: University of Texas at Austin.

Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic inquiry and word count*: LIWC2015 [Computer Software]. Austin, TX: Pennebaker Conglomerates.

Pennebaker, J., Mehl, M., & Niederhoffer, K. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*, 547-577. Retrieved January 14, 2018, from http://www.annualreviews.org.ezproxyd.bham.ac.uk/doi/pdf/10.1146/ annurev.psych.54.101601.145041

Plant, E. A., Hyde, J. S., Keltner, D., & Devine, P. G. (2000). The gender stereotyping of emotions. *Psychology of Women Quarterly*, 24(1), 81-92.

Plester, B., & Wood, C. (2009). Exploring relationships between traditional and new media literacies: British preteen texters at school. *Journal of Computer-Mediated Communication, 14*(4), 1108–1129.

Ponocny, I. (2001). Nonparametric goodness-of-fit tests for Rasch models. *Psychometrika, 66*(3), 437-460.

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

Prasath, R. R. (2010). Learning age and gender using co-occurrence of non-dictionary words from stylistic variation. *Rough Sets and Current Trends in Computing 7th International Conference, RSCTC 2010* (pp. 544-550). Warsaw: Springer-Verlag Berlin Heidelberg.

Prattala, R., Paalanen, L., Grinberga, D., Helasoja, V., Kasmel, A., & Petkeviciene, J. (2007). Gender differences in the consumption of meat, fruit and vegetables are similar in Finland and the Baltic countries. *The European Journal of Public Health, 17*(5), 520-525.

Punyanunt-Carter, N. M., & Hemby, C. O. (2006). College students' gender differences regarding e-mail. *College Student Journal, 40*(3), 651-653.

Qadeer, M. A. (2006). *Pakistan: Social and cultural transformations in a Muslim nation.* Oxfordshire, UK: Routledge.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language.* London: Longman.

Rafi, M. S. (2010). The sociolinguistics of SMS ways to identify gender boundaries. In R. Taiwo (Ed.), *Handbook of research on discourse behavior and digital communication: Language structures and social interaction.* Hershey, PA, USA: Information Science Reference.

Rafi, M. S. (2008). SMS text analysis: Language, gender and current practices. *Online Journal of TESOL France.* Retrieved November 3, 2017, from https://s3.amazonaws.com/ academia. edu.documents/32675446/SMS_Text_Analysis....pdf?AWSAccessKeyId=AKIAIWOWY YGZ2Y53UL3A&Expires=1509703685&Signature=up2ZF9XJMw22JMr%2FBxOYQX w40fg%3D&response-content-disposition=inline%3B%20filename%3DSMS_Text_Ana lysis_Langu

Rafi, M. S. (2017). Bilingualism and identity construction in the. *Journal of Multicultural Discourses, 12*(3), 254-271. Retrieved October 31, 2017, from http://dx.doi.org/10.1080/ 17447143.2017.1342649

Ragheb, M. G. (1980). Interrelationships among leisure participation, leisure satisfaction and leisure attitudes. *Journal of Leisure Research, 12*, 138-149.

Rahman, T. (1999). *Language, education and culture.* Karachi: Oxford University Press.

Rahman, T. (2010). *Linguistics for beginners: Basic concepts.* Karachi: Oxford University Press.

Ramirez-Esparza, N., Pennebaker, J. W., Garcia, F. A., & Suria, R. (2007). The psychology of word: A computer programme that analyses texts in Spanish. *Revista Mexicana de Psicologia, 24*(1), 85-99. Retrieved October 14, 2017, from http://www.redalyc.org/ articulo.oa?id=243020635010

Rao, D., Yarowsky, D., Shreevats, A., & Gupta, M. (2010). Classifying latent user attributes in Twitter. *Proceedings of the 2nd International Workshop on Search and Mining User-Generated Content* (pp. 37–44). Toronto, ON, Canada: ACM Digital Library.

Rasmussen, M. K., & Pidgeon, A. M. (2011). The direct and indirect benefits of dispositional mindfulness on self-esteem and social anxiety. *Anxiety, Stress & Coping, 24*(2), 227-233. doi:10.1080/10615806.2010.515681

Raumolin-Brunberg, H. (1998). Social factors and pronominal change in the seventeenth century: The Civil-War effect? In J. Fisiak & M.Krygier (Eds). *Advances in English Historical Linguistics.* (pp. 361-388). Berlin, Germany: Mouton de Gruyter.

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*(1), 133–152.

Razali, N., & Wah, Y. (2010). Power comparisons of some selected normality tests. *Proceedings of the Regional Conference on Statistical Sciences*, 126-138.

Reid, S. A., Keerie, N., & Palomares, N. A. (2003). Language, gender salience, and social influence. *Journal of Language and Social Psychology, 22*, 210-233.

Reid, T. B. (1956). Linguistics, structuralism and philology. *Archivum Linguisticum, 8*(1), 28-37.

Reigan, J. W., & Shute, V. (1993). Basic research on the pedagogy of automated instruction. In T. d. Jong, H. Spada, & D. M. Towne (Eds.), *The use of computer models for explication, analysis, and experiential learning* (pp. 121-132). New York: Springer-Verlag.

Reppen, R. (2010). Building a corpus: What are the key considerations? In A. O'Keeffe, & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 31-37). New York, NY, USA: Routledge.

Ridgeway, C., & Diekema, D. (1989). Dominance and collective hierarchy formation in male and female task groups. *American Sociological Review, 54*(1), 79-93. Retrieved January 9, 2018, from http://www.jstor.org.ezproxyd.bham.ac.uk/stable/pdf/2095663.pdf

Rilea, S., Roskos-Ewoldsen, B., & Boles, D. (2004). Sex differences in spatial ability: A lateralization of function approach. *Brain and Cognition, 56*, 332-43.

Roberts, W. (2002). Are animals stuck in time? *Psychological Bulletin, 128*, 473-489.

Rochon, E., Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (2000). Quantitative analysis of aphasic sentence production: Further development and new data. *Brain and Language, 72*, 193-218.

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). Retrieved January 7, 2018, from http://onlinelibrary.wiley.com.ezproxye.bham.ac.uk/doi/10.1111/j.1083-6101.1997.tb00074.x/full

Romaine, S. (2003). Variation in language and gender. In J. Holmes, & M. Meyerhof (Eds.), *The handbook of language and gender* (pp. 98-118). Maiden, MA, USA: Blackwell Publishing Ltd.

Rosen, J. (2004). *Your blog or mine?* Retrieved October 26, 2017, from http://www.nytimes.com: http://www.nytimes.com/2004/12/19/magazine/your-blog-or-mine.html

Rossetti, P. (1998). Gender differences in e-mail communication. *The Internet TESL Journal*, 4(7), 1-6. Retrieved December 8, 2017, from http://iteslj.org/Articles/Rossetti-GenderDif.html

Rubin, D., & Greene, K. (1992). Gender-typical style in written language. *Research in the Teaching of English, 26*(1), 7–40.

Ruble, T. (1983). Sex stereotypes: Issues of change in the 1970s. *Sex Roles, 9*, 397-403.

Rudman, L. (1998). Self-promotion as a risk factor for women: The costs and benefits of counterstereotypical impression management. *Journal of Personality and Social Psychology, 74*, 629–645.

Rustagi, M., Prasath, R. R., Goswani, S., & Sarkar, S. (2009). Learning age and gender of blogger from stylistic variation. *Third International Conference on Pattern Recognition and Machine Intelligence, PReMI 2009* (pp. 205-212). New Delhi: Springer-Verlag Berlin Heidelberg.

Rustam, R. (2010). The use of catchy words: A case study from Pakistan. *Language in India, 10*(6), 19-27. Retrieved November 4, 2017, from http://www.languageinindia.com/june2010/v10i6june2010.pdf

Sacks, H. (1992). *Lectures on conversations.* Oxford: Blackwell.

Sadker, M., & Sadker, D. (1985). Sexism in the schoolroom of the 80's. *Psychology Today, 19*(3), 54-57.

Saeed, A., Ghani, D. M., & Ramzan, M. (2011). Gender difference and L2 writing. *International Research Journal of Arts & Humanities (IRJAH), 39*(39), 29-40. Retrieved November 4, 2017, from http://sujo.usindh.edu.pk/index.php/IRJAH/article/view/1148/1064

Saeed, A., Hussain, S., & Fatima, S. (2015). Prepositional errors in the writings of Pakistani ESL students. *International Journal of English Linguistics, 5*(3), 120-128.

Saenger, P. H. (1997). *Space between words: The origins of silent reading.* Stanford, California: Stanford University Press.

Saily, T., Siirtola, H., & Nevalainen, T. (2011). Variation in noun and pronoun frequencies in a sociohistorical corpus of English. *Literary and Linguistic Computing, 26*(2), 167-188. doi:10.1093/llc/fqr004

Saleem, F., & Zubair, S. (2013). (Under)representating women in curricula: A context analysis of Urdu and English textbooks at the primary level in Pakistan. *Pakistan Journal of Women's Studies: Alam-e-Niswan, 20*(1), 57-71. Retrieved from http://www. pakistanwomenstudies. com/abstracts/vol%2020%20no.%201/(UNDER)%20REPRESENTING%20WOMEN%20IN%20CURRICULA_.pdf

Savović, B., Ubović, M., & Radenović, S. (2018). Analysis of the motivational speech held by the National team coach before the final FIFA U-20 World Cup match. *Physical Culture, 72*(1). doi:10.5937/fizkul1801102s

Sanauddin, N. (2015). *Proverbs and patriarchy: Analysis of linguistic sexism and gender relations among the Pashtuns of Pakistan* (Doctoral dissertation). The College of Social Sciences, School of Social and Political Sciences. Glasgow: University of Glasgow, Scotland. Retrieved November 5, 2017, from http://theses.gla.ac.uk/6243/1/2015 SanauddinPhD.pdf

Sanders, B., Soares, M. P., & D'Aquila, J. M. (1982). The sex difference on one test of spatial visualization: A nontrivial difference. *Child Development, 53*, 1106-1109.

Sardinha, T. B. (1995). Intertextual lexical cohesion in newspaper reports. *The 40th Annual Conference of the International Linguistic Association, Georgetown.* Georgetown.

Scates, C. (1981). *A sociolinguistic study of male/female language in freshman composition.* University of Southern Mississippi, Hattiesburg, Mississippi.

Scheidt, L. A. (2009). *Diary weblogs as genre.* (Doctoral dissertation), Indiana University, School of Library and Information Science. Retrieved October 26, 2017, from http://www. professional-lurker.com/linked/2009/qp/Diary_Weblogs_as_Genre2009.pdf

Schler, J., Koppel, M., Argamon, S., & Pennebaker, J. (2005). Effects of age and gender on blogging. *American Association for Artificial Intelligence.* Retrieved March 18, 2018, from http://u.cs.biu.ac.il/~schlerj/schler_springsymp06.pdf

Schmader, T., Whitehead, J., & Wysocki, V. (2007). A linguistic comparison of letters of recommendation for male and female chemistry and biochemistry job applicants. *Sex Roles, 57*(7), 509-514.

Schmid, H. (2015). Does gender-related variation still have an effect, even when topic and (almost) everything else is controlled? In J. Daems, E. Zenner, K. Heylen, D. Speelman, & H.

Cuyckens (Eds.), *Change of paradigms – new paradoxes: Decontextualizing language and linguistics (Vol. 31)* (pp. 327-346). Berlin, Germany: De Gruyter Mouton.

Schmid, H.-J. (2003). Do women and men really live in different cultures? Evidence from the BNC. In B. Lewandowska-Tomaszczyk, & P. J. Melia (Eds.), *Corpus linguistics by the Lune: A festschrift for Geoffrey Leech* (pp. 185–221). Frankfurt: Peter Lang.

Schmitt, D. P. (2015). *Are women more emotional than men?* Retrieved January 10, 2018, from www.psychologytoday.com: https://www.psychologytoday.com/blog/sexualpersonalities/201504/are-women-more-emotional-men

Schultz, P. (2013). Gender variation in writing: Analyzing online dating ads. *Coyote Papers, 21.* Retrieved January 14, 2018, from http://hdl.handle.net/10150/271015

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS One, 8*(9), 1-16. Retrieved January 19, 2018, from http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0073791&type=printable

Seale, C. (2006). Gender accommodation in online cancer support groups. *Health, 10*(3), 345-360.

Seale, C., Ziebland, S., & Charteris-Black, J. (2006). Gender, cancer experience and internet use: A comparative keyword analysis of interviews and online cancer support groups. *Social Science & Medicine, 62*(10), 2577–2590. Retrieved October 17, 2017, from https://doi.org/10.1016/j.socscimed.2005.11.016

Seidlitz, L., & Diener, E. (1998). Sex differences in the recall of affective experiences. *Journal of Personality and Social Psychology, 74*, 262-271.

Seredin, P., & Lyell, J. (2017). Deception detection in Russian texts. *Proceedings of the Student Research Workshop at the 15th Conference of the European Chapter of the Association for Computational Linguistics,* (pp. 43–52). Valencia, Spain: Association for Computational Linguistics. Retrieved October 14, 2017, from https://aclanthology.info/pdf/E/E17/E17-4005.pdf

Severance, L. E. (2012). *The gender dynamics of dissent in organizations.* Faculty of the Graduate School. Maryland: University of Maryland.

Sexton, B. J., & Helmreich, R. L. (2000). Analyzing cockpit communications: The links between language, performance, and workload. *Human Performance in Extreme Environments, 5*(1), 63-68. Retrieved January 9, 2018, from http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1007&context=jhpee

Shah, S. F. (2012). Gender inclusion: A neglected aspect of the English textbooks in Pakistan. *International J. Soc. Sci. & Education, 3*(1).

Shah, S. K., Mahmood, R., & Qureshi, M. B. (2011). Gender based social constraints associated with language in Punjabi society. *Interdisciplinary Journal of Contemporary Research in Business, 3*(8), 118–159.

Shaikh, P. M., & Khan, U. (2012). Constructing gender identities in discourse: A critical discourse analysis of two short stories. *American International Journal of Contemporary Research, 2*(3), 153-160. Retrieved November 5, 2017, from http://www.aijcrnet.com/journals/Vol_2_No_3_March_2012/17.pdf

Shapiro, S., Wilk, M., & Chen, H. (1968). A comparative study of various tests for normality. *Journal of the American Statistical Association, 63*(324), 1343–1372.

Sheskin, D. J. (2000). *Handbook of parametric and nonparametric statistical procedures.* New York: Chapman & Hall/CRC.

Shofwan, H., Musli, A., & Sugeng, A. (2013). An analysis on the gender-based difference of the eight Year students' recount paragraph writing based on the grammatical errors at SMPN 3 Jnggawah Jember. *Jurnal Pendidikan dan Pembelajaran, 2*(2), 75-84.

Sinclair, J. (2004). Corpus and text: Basic principles. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 1-16). Oxford: Oxbow Books. Retrieved November 17, 2017, from http://ota.ox.ac.uk/documents/creating/dlc/

Sladek, R. M., Bond, M. J., & Phillips, P. A. (2010). Age and gender differences in preferences for rational and experiential thinking. *Personality and Individual Differences, 49*, 907–911. Retrieved January 8, 2018, from https://doi.org/10.1016/j.paid.2010.07.028

Soler, J. (2013). *Author profiling: Gender identification.* (Master's thesis), Universitat Pompeu Fabra, Barcelona, Spain.

Soriano, E. C., Rentscher, K. E., Rohrbaugh, M. J., & Mehl, M. R. (2017). Semantic corpus comparison analysis of couple-focused interventions for problematic alcohol use. *Clinical Psychology and Psychotherapy, 24*, 618–631. Retrieved November 23, 2017, from http://onlinelibrary.wiley.com/doi/10.1002/cpp.2030/epdf

Speer, S. A. (2005). *Gender talk: Feminism, discourse and conversation analysis.* London: Routledge.

Spender, D. (1980). *Man made language.* London: Routledge.

Squires, L. M. (2010). Enregistering internet language. *Language in Society, 39*(4), 457-492. doi:10.1017/S0047404510000412

Squires, L. M. (2007). What's the use of apostrophes?: Gender difference and linguistic variation in instant messaging. *American University TESOL Working Papers Number 4.* doi:https://doi.org/10.17606/M6DQ0G

Statistica. (2017). *Cumulative total of Tumblr blogs from May 2011 to October 2017 (in millions).* Retrieved October 26, 2017, from www.statistica.com: https://www.statista.com/statistics/256235/total-cumulative-number-of-tumblr-blogs/

Stipek, D. J., & Gralinski, J. H. (1991). Gender differences in children's achievement-related beliefs and emotional responses to success and failure in mathematics. *Journal of Educational Psychology, 83*(3), 361-371.

Stubbs, M. (1996). *Text and corpus linguistics.* Oxfordshire, UK: Blackwell Publishers.

Stubbs, M. (2010). Three concepts of keywords. In M. Bondi, & M. Scott (Eds.), *Keyness in texts* (pp. 21-42). Philadelphia, PA, USA: John Benjamins Publishing Company.

Suborn, F. (2013). Gender differences in the use of linguistic forms in the speech of men and women in the Malaysian context. *Journal of Humanities and Social Science, 13*(3), 67-79.

Subramaniam, L. V., Roy, S., Faruquie, T. A., & Negi, S. (2009). A survey of types of text noise and techniques to handle noisy text. *Proceedings of the Third Workshop on Analytics for Noisy Unstructured Text Data July 23-24, 2009* (pp. 115-122). Barecelona, Spain: Association for Computing Machinery. Retrieved November 15, 2017, from https://dl.acm.org/citation.cfm?id=1568315

Suddendorf, T., & Corballis, M. (1997). Mental time travel and the evolution of the human mind. *Genetic, Social and General Psychology Monographs, 123*, 133-167.

Sunderland, J. (2008). *Language and gender: An advanced resource book.* New York: Routledge.

Swacker, M. (1976). Women's verbal behavior at learned and professional conferences. In B. Dubois, & I. Crouch (Eds.), *The sociology of the languages of American women* (pp. 155-160). San Antonio: Trinity University.

Swann, J. (2002). Yes, but is it gender? In L. Litosseliti and J. Sunderland (Eds.). *Gender identity and discourse analysis* (pp. 43–67). Amsterdam: John Benjamin.

Talbot, M. M. (2010). *Language and gender.* Cambridge: Polity Press.

Tamres, L. K., Helgeson, V. S., & Janicki, D. (2002). Sex differences in coping behaviour: A meta-analytic review and an examination of relative coping. *Personality and Social Psychology Review, 1*, 2-30.

Tannen, D. (1986). *That's not what I meant! How conversational style makes or breaks your relations with others* (First ed.). New York, USA: William Morrow & Company Inc.

Tannen, D. (1990a). Gender differences in topical coherence: Creating involvement in best friends' talk. *Discourse Processes, 13*(1), 73-90. doi:10.1080/01638539009544747

Tannen, D. (1990b). *You just don't understand: Women and men in conversation.* New York: Ballantine Books.

Tannen, D. (1993). The relativity of linguistic strategies: Rethinking power and solidarity in gender and dominance. In Deborah T. (Ed.), *Gender and conversational interaction* (pp. 165–188). New York: Oxford University Press.

Tannen, D. (1994). *Gender and discourse.* New York: Oxford University Press.

Tannen, D. (2010). *"Can't we talk?" (condensed from: You Just Don't Understand).* Retrieved October 4, 2017, from Linguistics: https://linguisticus.wordpress.com/2010/03/14/you-just-dont-understand-by-deborah-tannen/

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24–54. Retrieved January 8, 2018, from http://journals.sagepub.com.ezproxyd.bham.ac.uk/doi/pdf/10.1177/0261927X09351676

Tavakoli, H. (2012). *A dictionary of research methodology and statistics in applied linguistics.* Tehran, Iran: Rahnama Press.

Taylor, K., Thorne, S., & Oliffe, J. L. (2015). It's a sentence, not a word: Insights from a keyword analysis in cancer communication. *Qualitative Health Research*, 110–121. doi:10.1177/1049732314549606 qhr.sagepub.com

The Regents of the University of California. (2007). *Ninth Emeritus lecture honoring John J. Gumperz.* Retrieved October 3, 2017, from Anthropology Emeritus Lecture Series at U.C. Berkeley: http://www.lib.berkeley.edu/ANTH/emeritus/gumperz/gumptalk.html

Thelwall, M. (2008). Fk yea I swear: Cursing and gender in MySpace. *Corpora, 3*(1), 83–107. doi:10.3366/E1749503208000087

Thelwall, M., & Wouters, P. (2005). What's the deal with the web/blogs/the next big technology: A key role for information science in e-social science research? In F. Crestani, & I. Ruthven (Ed.), *Context: Nature, impact, and role, proceedings of 5th international conference on conceptions of library and information sciences, CoLIS 2005 Glasgow, UK, June 4-8, 2005* (pp. 187-199). Berlin Heidelberg: Springer. Retrieved October 25, 2017, from https://doi.org/10.1007/11495222_15

Thibaut, J. W., & Kelley, H. H. (1959). *The social psychology of groups.* New York, NY, USA: Wiley.

Thode, H. J. (2002). *Testing for normality.* New York: Marcel Dekker.

Thomson, R., & Murachver, T. (2001). Predicting gender from electronic discourse. *British Journal of Social Psychology, 40*, 193–208.

Thompson, B. L., & J. W. (2007). Everyday mindfulness and mindfulness meditation: Overlapping constructs or not? *Personality and Individual Differences, 43*, 1875–1885.

Thorne, B., & Henley, N. (1975). Difference and dominance: An overview of language, gender and society. In M. Rowley (Ed.), *Language and sex: Difference and dominance* (pp. 5-42). New York: Newbury House Publisher.

Thurlow, C. (2006). From statistical panic to moral panic: The metadiscursive construction and polar exaggeration of new media language in the print media. *Journal of Computer-Mediated Communication, 11*(3), 165-188.

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

Treem, J. W., & Thomas, K. Y. (2010). What makes a blog a blog? Exploring user conceptualizations of an old "new" online medium. *Proceedings of the Fourth International Conference on Weblogs and Social Media, 23–26 May*, (pp. 347-350). Washington, DC. Retrieved October 26, 2017, from http://www.aaai.org/Library/ICWSM/ icwsm10contents.php

Troll, L., Miller, S., & Atchley, R. (1979). *Families in later life.* California: Wadsworth.

Trudgill, P. (1972). Sex, covert prestige and linguistic change in urban British English of Norwich. *Language in Society, 1*(2), 179-195. Retrieved October 9, 2017, from http://www.jstor.org/ stable/ 4166683

Trudgill, P. (1977). The social differentiation of English in Norwich. In N. Coupland, & A. Jaworskin (Eds.), *Sociolinguistics: A reader* (pp. 179-183). New York: St. Martin's Press.

Trudgill, P. (2000). *Sociolinguistics: An introduction to language and society* (4th ed.). London: Penguin.

Ullah, H., Khan, A. N., Khan, H. N., & Ibrahim, A. (2016). Gender representation in Pakistani print media: A critical analysis. *Pakistan Journal of Gender Studies, 12*, 53-70. Retrieved November 5, 2017, from http://eprints.mdx.ac.uk/20246/

Umberson, D. (1992). Gender, marital status and the social control of health behavior. *Social Science & Medicine, 34*(8), 907-917.

Vali, S., & Kianiparsa, P. (2010). Gender differences in using cohesive devices while writing. *ELT Weekly*, n.p. Retrieved February 22, 2017, from http://eltweekly.com/2010/07/66-research-paper-gender-differences-in-using-cohesive-devices-while-writing-by-sara-vali-and-parnaz-kianiparsa/

Waldron, I. (1988). Gender and health-related behavior. In G. D. S (Ed.), *Healrh behavior: Emerging research perspectives* (pp. 193-208). New York: Plenum Press.

Wang, M. R., & Hsieh, S. C. (2007). Gender differences in the language for emotions. *Asian Journal of Management and Humanity Science, 2*(1), 89-97.

Wardhaugh, R. (2006). *An introduction to sociolinguistics* (5th ed.). Oxford, UK: Blackwell Publishing.

Watts, R. J. (2003). *Politeness.* Cambridge, UK: Cambridge University Press.

Watt, D. (2007). Variation and the variable. In C. Llamas, L. Mullany, & P. Stockwell (Eds.), *The Routledge companion to sociolinguistics* (pp. 3-11). Oxford: Routledge.

West, C., & Zimmerman, D. (1983). Small insults: A study of interruptions in cross-sex. In B. Thorne, C. Kramarae, & N. Henley (Eds.), *Language, gender and society* (pp. 102-117). Rowley, MA: Newbury House.

West, C., & Zimmerman, D. (1987). Doing gender. *Gender and Society, 1*(2), 125–151. Retrieved October 5, 2017, from https://www.gla.ac.uk/0t4/crcees/files/summerschool/readings/WestZimmerman_1987_DoingGender.pdf

Eakins, B. W., & Eakins, R. G. (1978). *Sex Differences in Human Communication.* Boston: Houghton Mifflin.

Widdowson, H. G. (1998). Context, community, and authentic language. *TESOL Quarterly, 32*(4), 705-716. Retrieved 11 20, 2017, from http://www.jstor.org/stable/3588001

Wiley, M., & Eskilson, A. (1985). Speech style, gender stereotypes, and corporate success: What if women talk more like men? *Sex Roles, 12*, 993-100.

Williams, G. (2002). In search of representativity in specialised corpora: Categorisation through collocation. *International Journal of Corpus Linguistics, 7*(1), 43-64. doi: https://doi.org/10.1075/ijcl.7.1.03wil

Winn, L., & Rubin, D. (2001). Enacting gender identity in written discourse: Responding to gender role bidding in personal ads. *Journal of Language and Social Psychology, 20*(4), 393–418.

Wissen, L. v., & Boot, P. (2017). An electronic translation of the LIWC dictionary into Dutch. In I. Kosem, C. Tiberius, M. Jakubíček, J. Kallas, S. Krek, & V. Baisa (Ed.), *Electronic*

lexicography in the 21st century. Proceedings of eLex 2017 Conference. (pp. 703-715). Leiden, the Netherlands, 19–21 September 2017: Lexical Computing CZ s.r.o., Brno, Czech Republic. Retrieved January 01, 2018, from https://elex.link/elex2017/proceedings/eLex_2017_Proceedings.pdf

Witmer, D.F., & Katzman, S. (1997). Online smiles: Does gender make a difference in the use of graphic accents. *Journal of Computer-Mediated Communication, 2*(4), 1083-6101

Wolf, M., Horn, A., Mehl, M. R., Haug, S., Pennebaker, J., & Kordy, H. (2008). Computer-aided quantitative text analysis. *Diagnostica, 54*(2), 85-98. doi:10.1026 / 0012-1924.54.2.85

Wong, C., Amini, R., & Koninck, J. D. (2016). Automatic gender detection of dream reports: A promising approach. *Consciousness and Cognition, 44*, 20-28. Retrieved from http://dx.doi.org/10.1016/j.concog.2016.06.004

World Health Organisation. (2016). *World Health Organization – Diabetes country profiles, 2016.* World Health Organisation. Retrieved May 28, 2018, from http://www.who.int/diabetes/country-profiles/pak_en.pdf?ua=1

Worldometers. (2017, October 26). *Blog posts written today.* Retrieved October 26, 2017, from http://www.worldometers.info: http://www.worldometers.info/blogs/

Xiao, R. (2010). Corpus Creation. In N. Indurkhya, & F. Damerau (Eds.), *The handbook of natural language processing* (pp. 147-165). London: CRC Press.

Xiufang, X. (2013). Gender differences in using language. *Theory and Practice in Language Studies, 3*(8), 1485-1489.

Xu, W., Li, L., Stefanone, M., & Fu, Y. (2014). Does social media users' commenting behavior differ by their local community tie? A computer-assisted linguistic analysis approach. *First Monday, 19*(1).

Yaefer-Dror, M., Hall-Lew, L., & Deckert, S. (2002). It's not or isn't it? Using large corpora to determine the influences on contraction strategies. *Language Variation and Change*, 79–118. Retrieved January 17, 2018, from http://malcah.faculty.arizona.edu/sites/malcah.faculty.arizona.edu/files/mydetal02.lvc14.pdf

Yahr, E. (2008). *Search Archive.* Retrieved October 28, 2017, from www.ajrarchive.org:http://www. ajrarchive.org/article.asp?id=4464

Yale, R. N. (2007). *Instant messaging communication: A quantitative linguistic analysis* (Unpublished Mater's Dissertation). Miami University, Department of Communication, Oxford.

Yap, B., & Sim, C. (2011). Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation, 81*(12), 2141-2155.

Youngs, G. (2004). Cyberspace: The new feminist frontier. In K. Ross, & C. M. Byerly (Eds.), *Women and media: International perspectives* (pp. 185–208). Malden, MA, USA: Wiley-Blackwell.

Yu, B. (2014). Language and gender in congressional speech. *Literary and Linguistic Computing, 29*(1), 118-132. doi:10.1093/llc/fqs073

Yusuf, H. (2013). *Mapping digital media: Pakistan.* Open Society Foundations, Open Society Media Program. London: Open Society Foundations. Retrieved October 29, 2017, from https://www.opensocietyfoundations.org/sites/default/files/mapping-digital-media-pakistan-20130902.pdf

Zaheer, Z. (2017). Word formation process in SMS language: A prognosis of gender identity. *Linguistics & Literature Review, 3*(1). Retrieved November 4, 2017, from http://admin.journals.umt.edu.pk/Media/Site/journals/SubSites/llr/FileManager/Volume3Issue1/Zainabzaheer5.pdf

Zaidi, S. B. (2016). Gender roles in Pakistani-Urdu wedding songs. *Journal of English Education and Linguistic Studies, 3*(1). Retrieved November 5, 2017, from http://jurnal.stainkediri.ac.id/index.php/jeels/article/view/171

Zaini, A., Hazirah, A., Saadiyah, D., & Kemboja, I. (2012). Gender differences in the language use of Malaysian teen bloggers. *Journal of Language Studies, 12*(1), 105-124.

Zalesky, Z. (1994). *Psychology of future orientation.* Poland: Lublin University Press.

Zhang, C., & Zhang, P. (2010). *Predicting gender from blog posts.* Amherst: University of Massachusetts. Retrieved January 11, 2018, from https://web.stanford.edu/~pyzhang/papers/gender_prediction.pdf

Zhang, Y., Dang, Y., & Chen, H. (2011). Gender classification for web forums. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, 41*(4), 668-678. Retrieved October 20, 2017. doi: 10.1109/TSMCA.2010.2093886

Zijlstra, H., van Middendorp, H., van Meerveld, T., & Geenen, R. (2005). aliditeit van de Nederlandse versie van de Linguistic Inquiry and Word Count (LIWC). *Netherlands Journal of Psychology.* Retrieved October 14, 2017, from https://doi.org/10.1007/BF03062342

Zimbardo, P., & Boyd, N. (1999). Putting time in perspective: A valid, reliable, individual-differences metric. *Journal of Personality and Social Psychology, 17*(6), 1271-1288.

Zimbardo, P., & Boyd, N. (2008). *The time paradox: The new psychology of time that will change your life.* London: Rider.

Zimmerman, D. H., & West, C. (1975). Sex roles, interruptions and silences in conversation. In B. Thorne & N. Henley (Eds.), *Language and sex: Difference and dominance* (pp.105-129). Rowley, MA: Newbury House.

Zimmerman, J., Wolf, M., Bock, A., Peham, D., & Benecke, C. (2013). The way we refer to ourselves reflects how we relate to others: Associations between first-person pronoun use and interpersonal problems. *Journal of Research in Personality, 47*, 218-225. Retrieved January 9, 2018, from http://dx.doi.org/10.1016/j.jrp.2013.01.008

Zuckerman, M., & Kuhlman, D. M. (2000). Personality and risk-taking: common biosocial factors. *Journal of Personality, 68*, 999–1029.

Appendix A

LIWC2015 Output Variable Information

Category	Abbrev	Examples	Words in category	Internal Consistency (Uncorrected α)	Internal Consistency (Corrected α)
Word count	WC	-	-	-	-
Summary Language Variables					
Analytical thinking	Analytic	-	-	-	-
Clout	Clout	-	-	-	-
Authentic	Authentic	-	-	-	-
Emotional tone	Tone	-	-	-	-
Words/sentence	WPS	-	-	-	-
Words > 6 letters	Sixltr	-	-	-	-
Dictionary words	Dic	-	-	-	-
Linguistic Dimensions					
Total function words	funct	it, to, no, very	491	.05	.24
Total pronouns	pronoun	I, them, itself	153	.25	.67
Personal pronouns	ppron	I, them, her	93	.20	.61
1st pers singular	i	I, me, mine	24	.41	.81
1st pers plural	we	we, us, our	12	.43	.82
2nd person	you	you, your, thou	30	.28	.70
3rd pers singular	shehe	she, her, him	17	.49	.85
3rd pers plural	they	they, their, they'd	11	.37	.78
Impersonal pronouns	ipron	it, it's, those	59	.28	.71
Articles	article	a, an, the	3	.05	.23
Prepositions	prep	to, with, above	74	.04	.18
Auxiliary verbs	auxverb	am, will, have	141	.16	.54
Common Adverbs	adverb	very, really	140	.43	.82
Conjunctions	conj	and, but, whereas	43	.14	.50
Negations	negate	no, not, never	62	.29	.71
Other Grammar					
Common verbs	verb	eat, come, carry	1000	.05	.23
Common adjectives	adj	free, happy, long	764	.04	.19
Comparisons	compare	greater, best, after	317	.08	.35
Interrogatives	interrog	how, when, what	48	.18	.57
Numbers	number	second, thousand	36	.45	.83
Quantifiers	quant	few, many, much	77	.23	.64
Psychological Processes					
Affective processes	affect	happy, cried	1393	.18	.57
Positive emotion	posemo	love, nice, sweet	620	.23	.64
Negative emotion	negemo	hurt, ugly, nasty	744	.17	.55
Anxiety	anx	worried, fearful	116	.31	.73
Anger	anger	hate, kill, annoyed	230	.16	.53
Sadness	sad	crying, grief, sad	136	.28	.70
Social processes	social	mate, talk, they	756	.51	.86
Family	family	daughter, dad, aunt	118	.55	.88

Category	Abbrev	Examples	Words in category	Internal Consistency (Uncorrected α)	Internal Consistency (Corrected α)
Friends	friend	buddy, neighbor	95	.20	.60
Female references	female	girl, her, mom	124	.53	.87
Male references	male	boy, his, dad	116	.52	.87
Cognitive processes	cogproc	cause, know, ought	797	.65	.92
Insight	insight	think, know	259	.47	.84
Causation	cause	because, effect	135	.26	.67
Discrepancy	discrep	should, would	83	.34	.76
Tentative	tentat	maybe, perhaps	178	.44	.83
Certainty	certain	always, never	113	.31	.73
Differentiation	differ	hasn't, but, else	81	.38	.78
Perceptual processes	percept	look, heard, feeling	436	.17	.55
See	see	view, saw, seen	126	.46	.84
Hear	hear	listen, hearing	93	.27	.69
Feel	feel	feels, touch	128	.24	.65
Biological processes	bio	eat, blood, pain	748	.29	.71
Body	body	cheek, hands, spit	215	.52	.87
Health	health	clinic, flu, pill	294	.09	.37
Sexual	sexual	horny, love, incest	131	.37	.78
Ingestion	ingest	dish, eat, pizza	184	.67	.92
Drives	drives		1103	.39	.80
Affiliation	affiliation	ally, friend, social	248	.40	.80
Achievement	achieve	win, success, better	213	.41	.81
Power	power	superior, bully	518	.35	.76
Reward	reward	take, prize, benefit	120	.27	.69
Risk	risk	danger, doubt	103	.26	.68
Time orientations	TimeOrient				
Past focus	focuspast	ago, did, talked	341	.23	.64
Present focus	focuspresent	today, is, now	424	.24	.66
Future focus	focusfuture	may, will, soon	97	.26	.68
Relativity	relativ	area, bend, exit	974	.50	.86
Motion	motion	arrive, car, go	325	.36	.77
Space	space	down, in, thin	360	.45	.83
Time	time	end, until, season	310	.39	.79
Personal concerns					
Work	work	job, majors, xerox	444	.69	.93
Leisure	leisure	cook, chat, movie	296	.50	.86
Home	home	kitchen, landlord	100	.46	.83
Money	money	audit, cash, owe	226	.60	.90
Religion	relig	altar, church	174	.64	.91
Death	death	bury, coffin, kill	74	.39	.79
Informal language	informal		380	.46	.84
Swear words	swear	fuck, damn, shit	131	.45	.83
Netspeak	netspeak	btw, lol, thx	209	.42	.82
Assent	assent	agree, OK, yes	36	.10	.39
Nonfluencies	nonflu	er, hm, umm	19	.27	.69
Fillers	filler	Imean, youknow	14	.06	.27

Category	Abbrev	Examples	Words in category	Internal Consistency (Uncorrected α)	Internal Consistency (Corrected α)
Punctuation					
Total Punctuation	-	-	-	-	-
Period	-	-	-	-	-
Comma	-	-	-	-	-
Colons	-	-	-	-	-
Semicolons	-	-	-	-	-
Question marks	-	-	-	-	-
Exclamation marks	-	-	-	-	-
Dashes	-	-	-	-	-
Quotation marks	-	-	-	-	-
Apostrophes	-	-	-	-	-
Parentheses	-	-	-	-	-
Other punctuation	-	-	-	-	-

"Words in category" refers to the number of different dictionary words and stems that make up the variable category. All alphas were computed on a sample of ~181,000 text files from several of our language corpora (see Table 2). Uncorrected internal consistency alphas are based on Cronbach estimates; corrected alphas are based on Spearman Brown. See the Reliability and Validity section below. Note that the LIWC2015 dictionary generally arranges categories hierarchically. There are some exceptions to the hierarchy rules. For example, Social processes include a large group of words that denote social processes, including all non-first-person-singular personal pronouns as well as verbs that suggest human interaction (talking, sharing) — many of these words do not belong to any of the Social processes subcategories. Another example is Relativity, which includes a large number of words that cannot be found in any of its subcategories.

(Source: Reproduced verbatim from Pennebaker, Boyd, Jordan, & Blackburn, 2015, p.3-4)

Appendix B

Sample of LIWC2015 Results

Filename	Seg	WC	Analytic	Clout	Auth	Tone	WPS	Sixltr	Dic	func
1	1	335	93.12	65.11	12.27	17.14	23.93	27.76	78.51	48.66
2	1	2229	89.88	65.53	12.92	52.71	22.07	23.15	83.98	49.21
3	1	12216	81.47	61.10	34.99	43.03	24.09	22.62	82.12	51.26
4	1	361	95.91	54.42	46.66	57.19	25.79	24.65	80.61	50.97
5	1	968	82.84	83.69	17.94	54.95	26.16	18.70	86.16	49.79
6	1	372	95.29	61.64	32.52	71.22	23.25	25.00	81.18	47.31
7	1	685	70.42	70.05	9.52	41.61	18.51	26.57	85.26	45.26
8	1	15376	90.90	64.74	22.37	28.04	21.06	24.95	76.89	46.75
9	1	10102	95.05	62.73	33.08	52.11	25.77	18.82	76.07	47.61
10	1	1077	94.54	55.16	20.96	5.74	26.92	28.97	78.74	44.57
11	1	839	44.93	55.20	48.26	7.72	23.31	20.14	85.70	57.33
12	1	2609	85.96	60.44	20.16	56.92	21.04	24.07	79.99	51.40
13	1	640	90.44	66.91	10.39	20.99	19.39	26.25	79.84	49.53
14	1	1460	45.57	44.83	65.71	28.05	18.02	17.67	87.88	56.64
15	1	1262	81.85	74.98	12.26	46.19	16.83	23.22	77.42	50.00
16	1	653	97.25	57.89	54.93	13.36	28.39	29.25	75.04	39.82
17	1	1811	75.79	79.93	42.23	6.27	18.67	18.66	84.59	50.14
18	1	1970	78.95	51.02	64.94	42.34	25.92	21.68	87.92	56.24
19	1	2160	74.57	68.81	24.46	54.53	18.46	20.65	86.53	52.50
20	1	568	35.49	33.62	54.35	92.81	19.59	24.65	90.67	58.10
21	1	1342	69.75	77.86	13.23	14.46	20.03	21.31	82.49	50.60
22	1	8490	92.59	68.36	22.14	32.23	26.95	20.82	78.92	50.65
23	1	4717	69.62	86.03	21.37	11.70	21.25	17.19	85.90	54.12
24	1	5844	87.99	69.93	27.07	39.21	17.09	19.13	83.68	53.35
25	1	656	84.83	54.84	34.43	89.52	19.29	25.91	85.52	51.98
26	1	824	84.05	69.06	22.08	15.52	28.41	24.64	76.70	46.72
27	1	965	82.04	64.56	48.86	74.08	20.53	20.52	80.10	47.46
28	1	10565	85.45	67.38	10.12	31.42	24.57	20.93	75.72	48.03
29	1	758	94.56	58.35	13.97	7.49	30.32	25.59	73.75	43.67
30	1	2205	92.38	64.68	9.56	63.61	22.73	26.21	81.09	49.61
31	1	558	78.41	64.68	13.34	17.90	20.67	24.55	79.03	51.61
32	1	623	70.00	50.64	41.71	28.44	17.80	25.20	81.70	47.99
33	1	739	71.09	61.72	27.74	28.02	26.39	20.84	83.76	55.75
34	1	707	91.66	91.26	17.60	28.12	28.28	23.76	78.08	46.11

Note: Values under each language variable show the percentages of the occurrence of that language feature in a file. File names have been replaced with serial number to ensure anonymity of bloggers. Names of some variables have been abbreviated to fit page size.

Appendix C

Mann-Whitney U-Test Results

S No	LIWC Category	Mann-Whitney-Wilcoxon (MWW) Two-Tailed Test				
		U	U (stand)	Exp value	Var (U)	p value (U test)
1	Word Count	1034983	0.930	1014444	488116326	0.3526
Summary Language Variables						
2	Analytical thinking	1324955	14.054	1014444	488116474	<0.0001
3	Clout	834572	-8.141	1014444	488116484	<0.0001
4	Authentic	889487	-5.656	1014444	488116499	<0.0001
5	Emotional tone	1082953	3.101	1014444	488113270	0.0019
6	Word/sentence	1175315	7.281	1014444	488115973	<0.0001
7	Words > 6 letters	1239280	10.177	1014444	488116027	<0.0001
8	Dictionary words	658668	-16.103	1014444	488116143	<0.0001
Linguistic Dimensions						
9	Total function words	773401	-10.910	1014444	488115890	<0.0001
10	Total pronouns	694959	-14.461	1014444	488115793	<0.0001
11	Personal pronouns	696183	-14.405	1014444	488115655	<0.0001
12	1st person singular	728652	-12.941	1014444	487688598	<0.0001
13	1st person plural	929381	-3.851	1014444	488007734	<0.0001
14	2nd person	748314	-12.103	1014444	483534234	<0.0001
15	3rd pers singular	969408	-2.040	1014444	487386102	0.0414
16	3rd pers plural	939218	-3.405	1014444	488097224	0.0007
17	Impersonal pronouns	892223	-5.532	1014444	488111498	<0.0001
18	Articles	1296212	12.754	1014444	488113859	<0.0001
19	Prepositions	1127192	5.103	1014444	488113023	<0.0001

Appendix C (Continued)

Mann-Whitney U-Test Results

S No	LIWC Category	Mann-Whitney-Wilcoxon (MWW) Two-Tailed Test				
		U	U (stand)	Exp value	Var (U)	p value (U test)
20	Auxiliary verbs	930116	-3.817	1014444	488113345	<0.0001
21	Common Adverbs	810851	-9.215	1014444	488111060	<0.0001
22	Conjunctions	808636	-9.315	1014444	488110649	<0.0001
23	Negations	843487	-7.738	1014444	488099381	<0.0001
Other Grammar						
24	Common verbs	773889	-10.888	1014444	488115210	<0.0001
25	Common adjectives	1069407	2.488	1014444	488109264	0.0129
26	Comparisons	1106456	4.165	1014444	488102030	<0.0001
27	Interrogatives	819088	-8.842	1014444	488098890	<0.0001
28	Numbers	1196725	8.251	1014444	488106491	<0.0001
29	Quantifiers	966866	-2.154	1014444	488100465	0.0313
Psychological Processes						
30	Affective processes	904213	-4.989	1014444	488112689	<0.0001
31	Positive emotion	994652	-0.896	1014444	488110083	0.3703
32	Negative emotion	894530	-5.428	1014444	488110508	<0.0001
33	Anxiety	876509	-6.246	1014444	487750637	<0.0001
34	Anger	988045	-1.195	1014444	487972586	0.2321
35	Sadness	900056	-5.178	1014444	487941861	<0.0001
36	Social processes	684033	-14.955	1014444	488115746	<0.0001
37	Family	668868	-15.778	1014444	479689264	<0.0001

Appendix C (Continued)

Mann-Whitney U-Test Results

S No	LIWC Category	Mann-Whitney-Wilcoxon (MWW) Two-Tailed Test				
		U	U (stand)	Exp value	Var (U)	p value (U test)
38	Friends	868431	-6.654	1014444	481532250	<0.0001
39	Female references	564504	-20.562	1014444	478829417	<0.0001
40	Male references	1073662	2.681	1014444	487735057	0.0073
41	Cognitive processes	856382	-7.154	1014444	488114977	<0.0001
42	Insight	850199	-7.434	1014444	488104995	<0.0001
43	Causation	979447	-1.584	1014444	488100859	0.1132
44	Discrepancy	930409	-3.804	1014444	488098884	<0.0001
45	Tentative	890363	-5.616	1014444	488107366	<0.0001
46	Certainty	904558	-4.974	1014444	488095749	<0.0001
47	Differentiation	921413	-4.211	1014444	488108266	<0.0001
48	Perceptual processes	707022	-13.915	1014444	488107958	<0.0001
49	See	820156	-8.794	1014444	488086150	<0.0001
50	Hear	833550	-8.189	1014444	487915186	<0.0001
51	Feel	783448	-10.458	1014444	487860013	<0.0001
52	Biological processes	628454	-17.471	1014444	488108223	<0.0001
53	Body	763547	-11.358	1014444	487958161	<0.0001
54	Health	750971	-11.927	1014444	487968517	<0.0001
55	Sexual	894252	-6.046	1014444	395158844	<0.0001
56	Ingestion	821153	-8.805	1014444	481921970	<0.0001

Appendix C (Continued)

Mann-Whitney U-Test Results

S No	LIWC Category	Mann-Whitney-Wilcoxon (MWW) Two-Tailed Test				
		U	U (stand)	Exp value	Var (U)	p value (U test)
57	Drives	1124428	4.978	1014444	488114796	<0.0001
58	Affiliation	873112	-6.397	1014444	488111607	<0.0001
59	Achievement	1220849	9.342	1014444	488104260	<0.0001
60	Power	1259780	11.105	1014444	488113104	<0.0001
61	Reward	1060090	2.066	1014444	488093085	0.0388
62	Risk	1029075	0.662	1014444	488057896	0.5078
Time Orientation						
63	Past focus	1035377	0.947	1014444	488113984	0.3434
64	Present focus	807516	-9.366	1014444	488114893	<0.0001
65	Future focus	1053031	1.747	1014444	488086912	0.0807
66	Relativity	1095282	3.659	1014444	488115008	0.0003
67	Motion	981591	-1.487	1014444	488098858	0.1370
68	Space	1147585	6.026	1014444	488113308	<0.0001
69	Time	1016337	0.086	1014444	488112351	0.9318
Personal Concerns						
70	Work	1264873	11.335	1014444	488113427	<0.0001
71	Leisure	1039740	1.145	1014444	488098482	0.2522
72	Home	753113	-11.846	1014444	486690517	<0.0001
73	Money	1117463	4.664	1014444	487968165	<0.0001
74	Religion	1037167	1.030	1014444	486825519	0.3031
75	Death	1015452	0.046	1014444	480221112	0.9633

Appendix C (Continued)

Mann-Whitney U-Test Results

S No	LIWC Category	Mann-Whitney-Wilcoxon (MWW) Two-Tailed Test				
		U	U (stand)	Exp value	Var (U)	p value (U test)
Informal language						
76	Informal language	944802	-3.156	1014444	487003078	0.0016
77	Swear words	1011998	-0.138	1014444	313875837	0.8902
78	Netspeak	995748	-0.928	1014444	405948125	0.3534
79	Assent	933545	-3.810	1014444	450962889	0.0001
80	Nonfluencies	990602	-1.094	1014444	474645543	0.2738
81	Fillers	987988	-2.203	1014444	144214615	0.0276
Punctuation						
82	Total Punctuation	849797	-7.452	1014444	488115292	0.0001
83	Period	939521	-3.391	1014444	488108468	0.0007
84	Comma	931904	-3.736	1014444	488112156	0.0002
85	Colons	1008417	-0.275	1014444	480513515	0.7834
86	Semicolons	923248	-4.162	1014444	480190653	<0.0001
87	Question marks	862024	-6.914	1014444	485944377	<0.0001
88	Exclamation marks	882402	-6.220	1014444	450718140	<0.0001
89	Dashes	1120769	4.813	1014444	488048415	<0.0001
90	Quotation marks	820751	-8.772	1014444	487547626	<0.0001
91	Apostrophes	886262	-5.802	1014444	488104005	<0.0001
92	Parentheses	1078683	2.916	1014444	485220958	0.0035
93	Other punctuation	1027029	0.599	1014444	441653043	0.5493