'He blogged vs She blogged': A Corpus-Based Language Analysis of 11,000 Blog Posts

Ibrar Hussain Khan[1]
Ayaz Afsar[2]
Muhammad Amjad[3]

Abstract

This article aims at analysing differences in language use by men and women bloggers of Pakistani English e-newspapers. The corpus-based analysis compares two gendered corpora comprising 11258 blog posts in all (6706 posts from 1674 men and 4552 from 1212 women bloggers) collected from blog archives of leading English e-newspapers of Pakistan. The data comprised blog posts uploaded during November 01, 2008 to August 31, 2015. Using automated text analysis tools—Linguistic Inquiry & Word Count (LIWC) and AntConc—the study employed a quantitative top-down approach for analysing both the corpora along the 93 in-built language variables of LIWC2015. With XLSTAT, the LIWC-computed results were subsequently subjected to statistical analyses through data normality measures, two-tailed hypothesis tests, Bonferroni's Correction, and effect size calculations. As a result, several gender differences in language use were found, some consistent with the previous research conducted in the contexts outside Pakistan.

1. Introduction

Within human language one sometimes hears the expressions "female language" and "male language" (Karlsson, 2007, p.4). Based on many differences found in men and women, including their language use, two prominent theories have emerged, i.e. the *biological* and the *social constructionist*. The former describes gender in terms of biological sex assuming that men outsize and outpower women (Bergvall, 1999; Tannen, 1993); that gender polarities exist in language use; and that gender roles are static and contextually independent. The latter defines gender in the light of social contexts assuming that gender roles are fluid and contextually situated; that gendered identities are voluntary; and that males and

[1] PhD scholar, Department of English, International Islamic University, Islamabad
[2] Professor, Department of English, International Islamic University, Islamabad
[3] PhD scholar, Department of English, International Islamic University, Islamabad

females choose their gendered identities in particular situations (Leaper & Smith, 2004).

1.1 Statement of the Problem

The *social constructionist theory*—the focus of interest for many linguists now-a-days—views gender as socially and culturally constructed "something that is accomplished every time we speak" (Coates, 2004, p.7). This "fluid" (Speer, 2005, p.13) approach to language and gender challenges the *essentialist* view based on innate biological and psychological features that differentiate men and women. A preliminary look at the texts posted by the bloggers of English e-newspapers of Pakistan reveal the fact that there is a significant difference in the language used by men and women bloggers. A systematic and rigorous analysis, however, was needed to determine whether or not the language of these gendered blogs supports the *essentialist* view to gender differences in language use.

1.2 Hypotheses

The study is intended to analyse the written language of men and women bloggers of English e-newspapers in Pakistan with 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.

2. Literature Review

As an outcome of research on studying gender differences in language use, three influential works, i.e. Lakoff, (1975), Key (1975) along with Thorne and Henley (1975) emerged, giving birth to prominent theoretical approaches to the study of language and gender. Taking the risk of oversimplification, four theoretical approaches, viz. *deficit*, *dominance*, *difference* and *dynamic*, can be identified. Among these, the first three were rooted in the *essentialist* view based on *biological theory* of gender polarity whereas the fourth one takes a *non-essentialist* view having its roots in the *social constructionist theory* of gender fluidity. The *deficit* approach views women as disadvantaged in language use, with their language not conforming to the dominant but implicit male norms of speech. Lakoff (1975), its founding proponent, suggests ten assumptions about what she felt characterises 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. The *deficit* approach is now "dated" (Coates, 2007, p. 62) in research studies. The *dominance* approach, associated

with Don Zimmerman and Candace West (1975), Pamela M. Fishman (1980) and Dale Spender (1980), views women as the oppressed gender and explains gendered language differences (Fishman, 1997) in terms of men's dominant and women's subordinate roles. The *difference* approach is often offered as an alternative to the *dominance* approach to explain language differences by men and women. Behaviour that was once 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 fourth and most recently developed approach to language and gender is the *dynamic* approach, which has its underpinnings in the *social constructionist* theory. It views gender identity as something socially constructed rather than a predetermined construct (Coates, 2007). What is significant from the point of view of linguistics is that this approach is in sharp contrast to the *essentialist* view of gender adopted by the early three models: *deficit*, *dominance*, and *difference*. The basic contention between the *essentialist* approaches (*deficit*, *dominance*, *difference*) and the *social constructionist* approach (*dynamic*) is that the former consider language as the product of gender while the latter takes gender as the product of language.

Much of the previous research on gender and language use conducted outside Pakistan has been thoroughly reviewed elsewhere (e.g. Newman, Groom, Handelman, & Pennebaker, 2008; Mulac, Bradac & Gibbon, 2001). The subsequent part, therefore, is dedicated to a brief overview of what has been conducted in the context of Pakistan.

The linguistic landscape of Pakistan is rich as there are about 72 languages spoken in Pakistan (Rahman, 2010). Of these, English is one of the official languages in the country alongside Urdu. The rich interplay of these languages has its manifestations in spoken and written gendered language use in *bilingual* (Urdu-English) and *monolingual* contexts, both offline and online. Consequently, there is a growing body of research on gender and language in Pakistan. This limited review, however, restricts itself to research which was conducted on English language for exploring gender differences in language use both in intra- and inter-gender communication contexts.

A study on intra-gender conversation (Gul, 2010) described gender-based speech variations in the working environment of the Pakistan Air Force. While the study confirmed many of the previous research findings, it refuted others. In inter-gender conversation analysis, Noreen and Zubair (2012) pinned their study in the *non-essentialist* approach and analysed leisure talk recordings of eight conversations between close female friends for linguistic patterns of overlapping, tags, hedges, intensifiers, compliments, repetition, latching, and backchannel support. 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.

While these studies analysed spoken language, others examined written language. For example, a study of SMS writing by men and women (Rafi, 2008) randomly collected 100 text messages from 20 phones. The study analysed the selected texts at lexical, morphological and syntactic levels and concluded that there was a significant difference between males and females in lexical and morpho-syntactical choices. It also found that females were more skilful in writing longer, more complex and lexically denser. Later, the same researcher (Rafi, 2010) investigated gendered SMS texts against the variables of compression, abbreviations, symbols, tenses and punctuation to find out gender boundaries. He concluded that SMS text identifies some gender boundaries: for instance, females tend to use more compressed forms of words, abbreviations, and acronyms than males. Furthermore, he found significant differences in use of standard grammatical structures, and punctuation. Similarly, Rustam (2010) conducted a study on the use of 15 selected catchy words (e.g. 'chill pill', 'lash pash', 'fit fat') in a mix of spoken and written texts. His results showed that female students tended to use catchy words more than the male students. He also reported that catchy words were used more in SMS compared to their use in face-to-face communication, letters, and telephone calls. A more recent attempt on exploring gender differences in language of SMS was that of Zaheer (2017) who explored gender difference in word formation processes (WFPs) that are used in SMS language. She gathered data from 50 male and 50 female graduate students. She found that males and females follow different WFPs, both standard and non-standard. She also reported that females tend to use more WFPs like acronyms, clippings, homophones in comparison to males and that females do not use standard forms of language in their SMS writing. Based on these findings, she concluded that WFPs can mark gender identities in SMS.

In online *computer-mediated communication* (CMC), unfortunately, scarce research is available in Pakistan on gender differences in language use. Previous research has pre-dominantly focused on the analyses of language use on the Facebook. For instance, Nazir (2012) traced similarities and differences between men and women in 60 Facebook profiles (30 each for men and women). He found that women go for groups that relate to studies and fashion while men prefer groups that are related to business. Moreover, in choice of topics, he noted, women wrote more about their emotions, examinations, psychology, studies, fashion, and weather. In contrast, men talked more about business, politics and sports. Women were also found to use more emoticons to maintain face and keep conversation going. The study, however, did not find any significant difference in

the use of internet language by the two genders. In the same vein, Naveed's (2014) corpus-based analysis of gender boundaries in cyber language among 100 Facebook users of various age groups (13 to 30 years) revealed significant linguistic differences between males and females. A latest study by Kamran and Mansoor (2017) compared the written comments on Facebook by 220 Pakistani male (62%) and female (38%) university students. The study highlighted that female students used more emotional language in their comments and tended to post more comments about fashion, women's progress, diet, nostalgia and party arrangements.

An analysis of these studies shows that research on language differences between males and females is a growing phenomenon in Pakistan. However, most of the studies have analysed small language samples and applied traditional manual-coding techniques, which are not always consistent across studies. Moreover, a significant CMC area so far underrepresented in previous research is Pakistani blogosphere. The existing research on Pakistani blogs is not only limited in scope but also carried out with different research aims such as examining female gender portrayal in newspaper blogs (e.g. Amjad & Rasul, 2017) and developing a new computer algorithm for sentence-based semantic analysis of blogs (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. 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 such as *Daily Jang* and English dailies like *Dawn*, *The Daily Times*, *The Express Tribune*, *The Nation*, and *The News* publish blogs. The current study aims to add to the existing body of research by applying empirical corpus linguistics (CL) research techniques to analyse a large data set comprising more than 11,000 blog posts of Pakistani English e-newspapers.

3. Methodology

The study applied corpus-based quantitative approach to analysing two gendered corpora of blogs. The process involved two major steps, viz. building the corpora and analysing the corpora.

3.1 Building the GenCorB

The study required building and analysing two comparable 'gendered corpora of bloggers' (GenCorB): one 'corpus of men bloggers' (MenCorB) and the other 'corpus of women bloggers' (WenCorB). The entire GenCorB was manually retrieved from blogs posted between November 01, 2008 to August 31, 2015. The corpora building process involved three phases. The first phase comprised manual collection and storage of all blog posts in MS Word file format. Each file was named under the blogger's name preceded by a mark for gender. In all, 1674 files

for men bloggers and 1212 files for female bloggers were retrieved and stored in MS Word format (.doc). A major reason for saving the files initially in MS Word format was to use the inbuilt features (i.e. *Spelling & Grammar*) of the programme for clearing noise from the data.

After retrieval and storage of data, the second phase was to clean the data of any noise, which 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 any kind of sources or internet links, footnotes, editor's notes, pictures and cartoons, embedded videos, and advertisements. The second kind was of *internal nature*, i.e. it seemed to be originated by the bloggers and involved typological or formatting errors, text underlining, insertion of bullets, and use of extra dashes as section separators within blog posts.

Once all the blog posts were cleaned of noise, the third phase began which involved preparation of the data for preliminary computation. Since manual conversion of all 2886 MS Word (.doc) files (1674 files of MenCorB and 1212 files of WenCorB) into Text Documents (.txt) was a time-consuming and laborious task, a custom-built software named "FileAttributesReader" was developed with the help of a software engineer. Completion of the three phases of data collection and preparation made the corpora building process rather slow, which took about one and a half years to complete.

3.1.1 The Question of Representativeness

The current study employed an *external criterion* (Flowerdew, 2004) to achieve *representativeness* of the MenCorB and WenCorB, which called for a definition of the target population—adult male and female Pakistani bloggers of leading English e-newspapers of Pakistan. Therefore, applying *purposive sampling*, blog posts representative of the population were included in the study, which required determination of population boundaries (Biber, 1993). For this purpose, the blogs' language (i.e. English), type (e-newspaper blogs) and country context (Pakistan) were kept in mind. Also, the bloggers' ages, nationalities and genders were taken into consideration. Although bloggers' profiles were a useful source for obtaining such metadata, the process was not as simple and straightforward as it appeared to be. For instance, information on age in some of the cases was not explicitly stated. So, it was either obtained from profiles where such explicit information was available or inferred from related information that implicitly indicated the bloggers' ages (e.g. information on profession or level of education). Since the

current study was of adult bloggers, 18 years was considered as age threshold in line with the provisions of *The Majority Act 1875* (Government of Pakistan, 2006) and bloggers younger than 18 years were not included in the study.

The second boundary of the target population was the bloggers' genders, which was determined by their names and pictures given in the bloggers' profile. In some cases, the names did not clearly indicate the bloggers' genders; so, the accompanying photographs were of great help. Also, English gendered pronouns (*he*/*she*) used in the bloggers' profiles were helpful in finding out the bloggers' genders. 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). Also excluded were blog posts jointly written by a male and a female blogger or bloggers' groups for the obvious reason of indefinite contribution to the text by either gender. Similarly, anonymous posts or posts using vague aliases and initials (with no bloggers' photographs) were not included in the data. Besides age and gender, nationality of the bloggers was another boundary 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 to see which text category to be included in the data. Broadly speaking, the text category selected for the current study belonged to a particular web genre i.e. *blog*. Within this broader category, the texts were selected from Pakistani e-newspaper blogosphere, a sub-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 provide a huge and diversified linguistic resource to researchers as bloggers from different professional backgrounds contribute their posts.

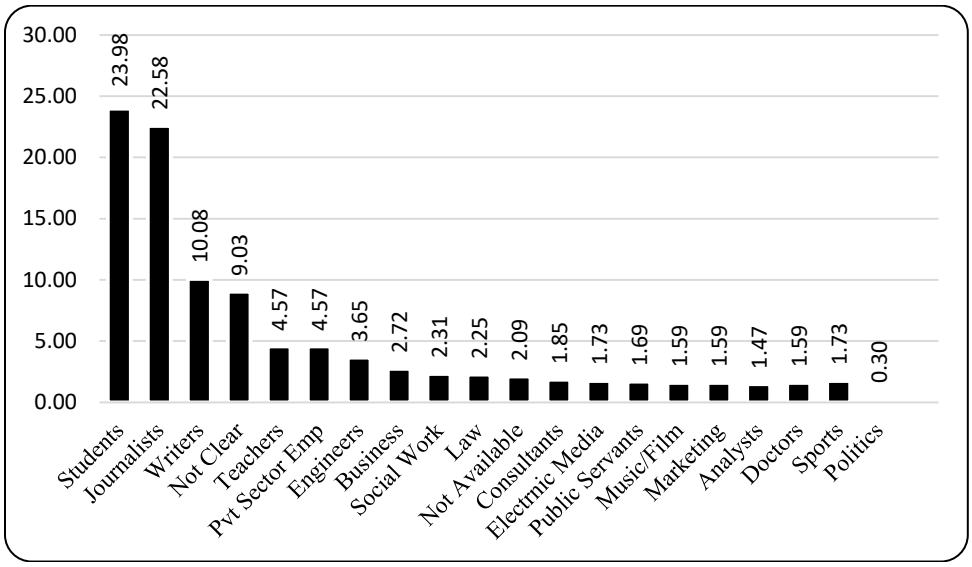


Figure 1. Percentage of bloggers' professions in GenCorB

Information on bloggers' professions was also available in a varied manner. 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). However, still others provided vague information or did not contain such information at all. Similarly, bloggers from some professions had a reasonable size in the population while others had lesser representation. The number of bloggers from homogenous professions with lesser representation was taken together to afford such professions visibility in data as shown in Figure 1, which details that information about professional background of the bloggers was available in 89% cases whereas in about 2% cases, the information was not available. In about 9% cases, the profession was not clearly mentioned.

3.1.2 Size of the GenCorB

Another important consideration in building the GenCorB was the corpora's size. Corpus size is a fluid concept with no ideal data volume. For determining corpus size, some studies (e.g. Gatto, 2014; Reppen, 2010) have recommended consideration of two major factors: the representation of the language being investigated and practicality. Others claim that, rather than corpus size, a thorough definition of the target population is the most important consideration in building corpora (Biber, 1993). In terms of tokenisation, Baker (2006) argues that for non-discourse oriented studies a million tokens 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 current study 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. The GenCorB comprised the whole texts (Sinclair, 2005) of all the posts of the Pakistani bloggers: the MenCorB contained about 5.6 million words whereas the WenCorB had around 3.3 million words. In total, 11258 blog posts (6706 for 1674 men and 4552 for 1212 women bloggers) were collected. The average length of post per blogger was 3386.80 tokens for MenCorB and 2801.30 tokens for WenCorB.

3.2 Analysing the GenCorB

Analysis of the GenCorB was carried out by two automated text analysis (ATA) tools: Linguistic Inquiry and Word Count (LIWC) 2015 and AntConc 3.4.4 (Anthony, 2016) together with a statistical package XLSTAT (Addinsoft, 2016). The analysis was completed in two stages. The first stage involved processing the data with LIWC2015 and XLSTAT (Addinsoft, 2016).

LIWC (pronounced "Luke") was initially developed as a result of factor analysis 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. LIWC is composed of two major parts: a processing feature and a dictionary. The former is the programme itself which opens and processes a number of text files at a time while the latter is a group of words that relate to each category (Tausczik & Pennebaker, 2010).

For its versatility, LIWC has also been adapted from English into other languages including Dutch (Zijlstra, van Middendorp, van Meerveld, & Geenen, 2005), Korean (Lee, Park, & Seo, 2006), Spanish (Ramirez-Esparza, Pennebaker, Garcia, & Suria, 2007), German (Wolf, et al., 2008), Chinese (Jinlan, et al., 2012), Catalan (Mass´o, Lambert, Rodr´ıguez, & Saur´ı, 2013), Italian, Portuguese (Filho, Pardo, & Alu´ısio, 2013), Norwegian (Johnsen, Vambheim, Wangberg, & C, 2014) and Russian (Seredin & Lyell 2017). Adaptation of this tool into some other languages is underway, including Turkish and Arabic. However, such adaptations have used LIWC2001 or LIWC2007 dictionaries rather than LIWC2015 (Pennebaker, Boyd, Jordan, & Blackburn, 2015).

LIWC2015 analyses the rate at which language is used and groups results under 93 categories. The tool processes each word in a text file by matching it to the 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, 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 processed, results for 93 categories in percentages were exported as MS Excel Worksheets by LIWC2015. The subsequent analysis of LIWC2015 results for obtaining inferential statistics, i.e. the hypothesis test, entailed working out means, standard deviations, effect sizes, data normality and p-values (two-tailed), which were calculated with the help of XLSTAT (Addinsoft, 2016)—a user-friendly statistical package, which easily integrates as an add-in feature into MS Excel Worksheet.

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). This widely applied ATA tool, besides offering other features, provides concordances, frequencies and collocations of any target expression. At this stage, selected expressions captured by LIWC2015 dictionaries were subjected to further analysis to see any gender differences in their use.

4. Results and Discussion

After initial descriptive statistics about MenCorB and WenCorB were obtained separately as percentages of language features across 93 LIW2015 variables and their means and standard deviations were calculated, an appropriate statistical measure was required to test the null hypothesis. The decision was made after testing data normality for all samples of MenCorB and WenCorB across 93 LIWC2015 variables through Jarque-Berra (JB) test— a goodness-of-fit measure for testing skewness and kurtosis matching a normal distribution. Since an analytic test like JB is not the only measure of data normality, the study also checked skewness and kurtosis through visual inspection of histograms, Probability-Probability (PP) plots and Quantile-Quantile (QQ) plots for each of the 93 LIWC2015 variables of both MenCorB and WenCorB. Since these tests indicated that the data in both the corpora tended to deviate from normal distribution, a nonparametric two independent sample test, i.e. Mann-Whitney U-test, was selected for testing the null hypothesis, which is used as an alternative to its parametric equivalent, i.e. the independent samples t-test. The level of risk, i.e. *alpha* (α), associated with the null hypothesis was set at 0.05.

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 93 LIWC2015 variables to test the null hypothesis. Therefore, another statistical measure was required to avoid family-wise Type I error. For this purpose, Bonferroni-Dunn

test (also called as Bonferroni's Correction) was conducted, which is computationally identical to multiple t-tests, Fisher's LSD or linear contrasts. It is calculated by the formula α/m where α is the desired alpha level and m represents the number of hypotheses being tested.

For Bonferroni-Dunn test, the highest family-wise alpha was stipulated to be 0.05, which was the alpha (α) set for each of the 93 LIWC2015 variables while conducing Mann-Whitney U-Test. That is to say that the study did not want to accept more than 5% chance of accepting Type I error on the whole set of comparisons against 93 LIWC2015 variables. For this purpose, the value of alpha (α) was divided by the total number of comparisons, i.e. 93 (the number of the LIWC2015) variables. The resulting more stringent value of α = α /93 = .0005376344 represented for each of the comparisons that were conducted across LIWC2015 variables the likelihood of committing Type I error. Simultaneously, once means and standard deviations for all the 93 LIWC variables were calculated, effect size for the study was computed with the help of Choen's (1988) *d*, which classifies effect sizes into *small* (.2), *medium* (.5) and *large* (.8) categories.

Table 1 combines the results of means, standard deviations, U-test and Cohen's *d* for the GenCorB. 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 1. Combined results: means, standard deviations, U-test and Cohen's d for the GenCorB

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 Summary Variables							
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
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
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	0.0003	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*
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*

As shown in Table 1, overall statistically significant gender difference in language use was found for 65 out of 93 language variables: Men scored higher on 15

whereas women overused 50 language categories. No statistically significant difference was found for the remaining 28 features.

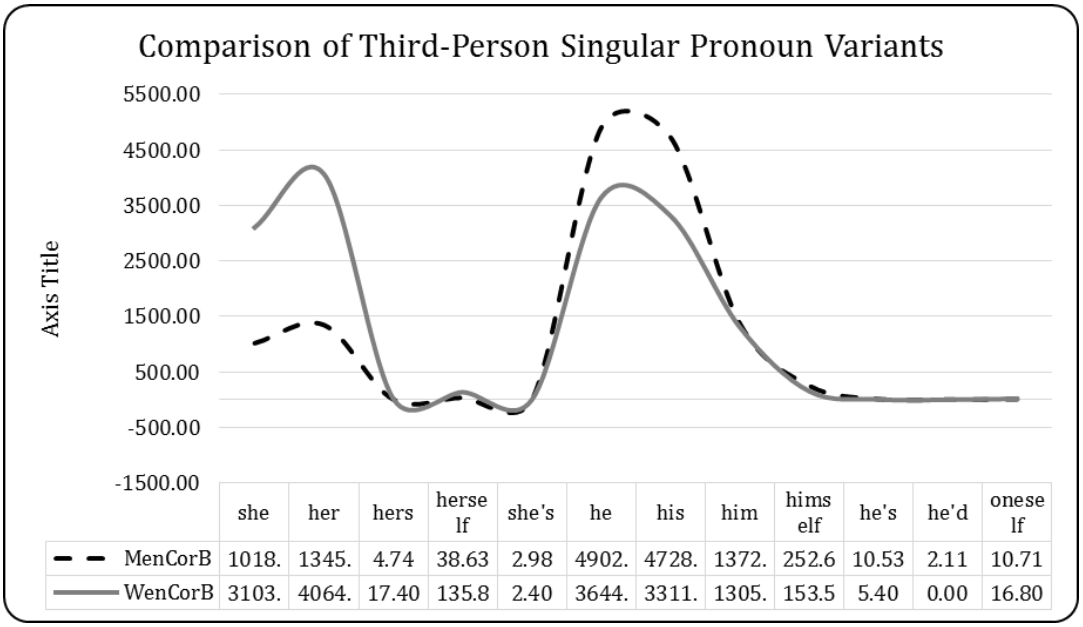
When results of the current research were compared to consistent language features used by men and women as reported by previous studies conducted outside Pakistan, some similarities in results surfaced, which merit attention. Of the 15 male-dominated language variables found by this study, results for men were consistent with the previous research in 06 language categories, i.e. overuse of *articles*, *numbers*, *words>6 letters* and expressions that relate to *analytical thinking*, *power*, and *space* (spatial ability). In contrast, of 50 female-dominated language variables found by this study, women were consistent with the previous research in the overuse of 16 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 sight*, *sense of touch*, *affiliation*, *home*, and *assent*.

In addition to the hypothesis test, the effect size of gender on language use was also calculated through Cohen's (1988) *d* as given in Table 1. It was observed that most of the effect sizes on the language dimensions analysed in this study were in the range of 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* scored between medium and large effect sizes. In men bloggers, *analytical thinking* crossed the threshold of medium effect sizes (.52). In fact, only five dimensions met Cohen's (1988) criterion for small effect—*long words*, *articles*, *swear words*, *social words* and *pronouns*. Differences along these five language dimensions, however, seem to be quite meaningful in gender and language use.

In interpreting the size of these effects, it is worth highlighting that 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 (1988) 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 from different walks of life. However, despite these quantitative differences in the use of language, it could not be claimed that men and women spoke 'different' languages. These differences could rather be seen across a quantitative continuum which ranges from no difference or similarity to large size differences.

Alongside conducting the hypothesis test and measuring gender effect sizes for 93 LIWC2015 variables, the data was further examined for any subtle gender differences in language use that occurred in particular expressions captured by LIWC2015 dictionaries. This part of analysis was conducted with AntConc 3.4.4 (Anthony, 2016) and results were compared to previous research. Due to space restrictions, examples are presented here for one language category, i.e. *third-person singular pronoun*.

Previous research conducted on studying gender differences in the use of third-person singular pronoun (*he* and *she*) has yielded mix results. Some studies have reported that women overuse third-person singular pronoun (Argamon, Koppel, Fine, & Simoni, 2006; Newman, Groom, Handelman, and Pennebaker, 2008; Friginal 2009; Yu, 2014) while others have highlighted additional aspects of gender differences in the use of personal third-person singular pronoun. For example, Herring and Paolillo (2006) concluded that while third-person masculine was favoured by women, third-person feminine did not show any significant gender differences. Similarly, Ahmad and Mehmood (2015) found that men overused masculine third-person singular (*he*) whereas women overused feminine third-person singular (*she*).



	she	her	hers	herself	she's	he	his	him	himself	he's	he'd	oneself
MenCorB	1018.	1345.	4.74	38.63	2.98	4902.	4728.	1372.	252.6	10.53	2.11	10.71
WenCorB	3103.	4064.	17.40	135.8	2.40	3644.	3311.	1305.	153.5	5.40	0.00	16.80

Figure 2. Comparison of personal third-person singular pronoun variants; frequency normalized to one million

Yet, some research has not found any significant differences in the overall use of third-person singular pronoun by men and women (e.g. Lenard, 2017). While the overall result of the current study for third-person singular pronoun (*he*/*she*) and its variants was consistent with Lenard (2017), it was observed, as shown in Figure 2, that both men and women bloggers overused their own gender-directed personal third-person singular pronoun. As a result, men tended to overuse third-person singular pronoun *he, his, him, himself* and women overused *she, her, hers,* and *herself*. This observation 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 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. 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 the grammatical gender sub-categories of nouns, i.e. masculine, feminine and neuter nouns, which yielded some interesting patterns as indicated by Figures 3 and 4.

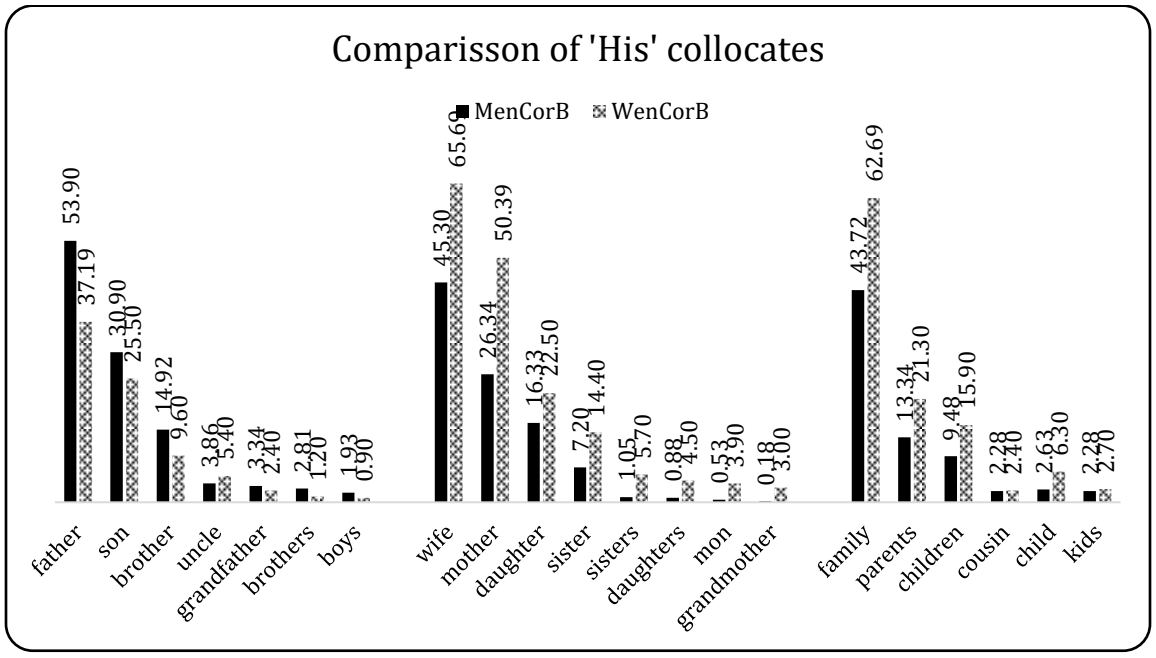


Figure 3. R1 noun-referents (masculine, feminine, neuter) to family members with 'His' in GenCorB; frequency normalized to one million

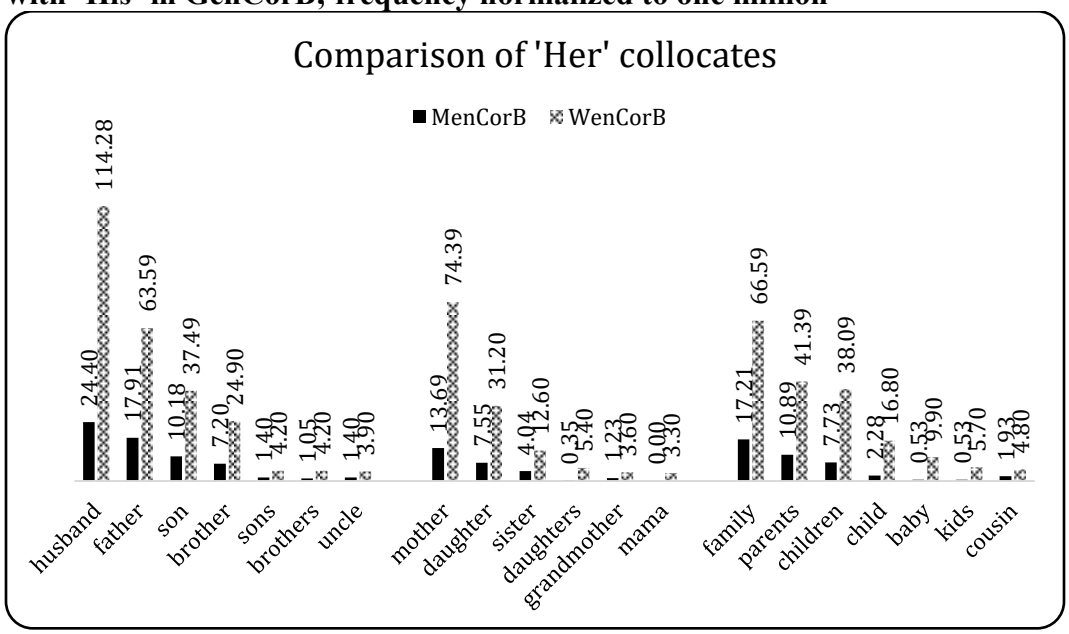


Figure 4. R1 noun-referents (masculine, feminine, neuter) to family members with 'Her' in GenCorB: frequency normalized to one million

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 or women that they referred to with *his*/*her*. 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, signals, gender polarity in men and women.

Another subtle difference was the collocation of the words *grandfather* with the third-person possessive pronoun *his*/*her*. Interestingly, 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).

Besides analysing frequencies, concordances for *his/her grandfather* and *his/her grandmother* were also generated and analysed to note any similarity of differences in their use as shown in Concordances 1 to 4 below. In these concordances, names of persons have been replaced with four asterisks (****) to ensure their anonymity.

'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

****. 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	***His***	would say that the morning

part of that. ***grandfather*** prayers would

'grandfather' as someone to be emulated

y, he likely feels compelled to deliver robust speeches like ***his grandfat*** ***her*** . 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 Hernandez, 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 ****, he forgot the obvious yet delicate

It seemed as if, in attempt to replicate ***his*** ***grandfather*** in his signature awami shalwar suit with

of Pakistan. Baby **** might dress like

'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 1 'his grandfather' occurrences in GenCorB

As shown in Concordance 1, 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*.

'grandfather' as ordinary family elderly relation

, where she spent most of her time, ***her*** ***grandfather*** playing his tabla. The 70-year-old dancing to 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 2 'her grandfather' occurrences in GenCorB

In contrast, as indicated by Concordance 2, in all occurrences of *her grandfather*, the relationship of grandfather was presented as an *ordinary family elder* in both the corpora. 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. Besides, the concordances of *his grandmother* and *her grandmother* were also studied and some subtle patterns were observed as indicated by Concordances 3 and 4.

'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 **** 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, ****, 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***	. **** 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.
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 3 '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

	her grandmother	
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

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

Concordance 4 'her grandmother' occurrences in GenCorB

It can be noted that for the words *grandfather* and *grandmother*, the concordances of third-person singular pronouns (*his*/*her*) presented a contrasting pattern. With third-person singular possessive cases occurring 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 wider connotations, i.e 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 in the use of the words *grandfather* and *grandmother* in another sense was also noted. 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 relate to beauty and looks.

5. Conclusion

Besides exploring gender differences in language use in Pakistani blogosphere, the current study aimed to verify the claims of the *essentialist* theory as contended by the *social constructionist* theory. The question it tried to answer was whether stereotypical language differences as theorised by the *essentialist* approach to gender differences found in Pakistani social and CMC contexts. The results indicated that out of the 93 variables analysed in the study, there were 65 language dimensions on which genders differed. In particular, consistency of the present findings with those of the previous studies for 21 language features besides uniformity between present results pertaining to gender effect sizes and those of the previous research conducted outside Pakistan make an indication to the fact that some universal gender behaviour patterns do exist as claimed by the *biological theory*, which also manifest themselves in language use. This, in turn, makes a case for the relevance of the claims of the *essentialist* approach to language and gender research.

References

Addinsoft (2016). *XLSTAT 2016: Data analysis and statistical solution for Microsoft Excel* [Computer Software]. Paris, France: Addinsoft.

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/

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?acco untid=8630

Anthony, L. (2016). *AntConc (Version 3.4.4)* [Computer Software]. Tokyo, Japan: Waseda University. Retrieved November 23, 2017, from http://www.laurenceanthony.net/

Argamon, S., Koppel, M., Fine, J., & Shimoni, A. R. (2006). 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

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

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

Bergvall, V.L. (1999). Toward a comprehensive theory of language and gender. *Language in Society, 28*, 273-293.

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

Coates, J. (2007). Gender. In C. Llamas, L. Mullany, & P. Stockwell (Eds.), *The Routledge companion to sociolinguistics* (pp. 62-68). Oxford: Routledge.

Coates, J. (2004). *Women, men and language* (3rd ed.). Harlow: Longman.

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.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Lawrence Erlbaum Associates, Publishers.

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.

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

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.

Fishman, P. (1980). Conversational insecurity. In G. Howard, R. Robinson, & P. Smith (Eds.), *Language: Social psychological perspectives* (pp. 127-132). New York: Pergamon Press.

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.

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.

Government of Pakistan. (2006). *The majority act, 1875* (6th ed.). Lahore: Legal Research Centre.

Gul, S. (2010). *Studying gender related views about speech variations at Pakistan Air Force*: *A case study* (Unpublished M Phil thesis). IIU, Islamabad.

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

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

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

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

Karlsson, S. (2007). *Gender-related differences in language use* (Unpublished Master's thesis). Luleå University of Technology, Luleå, Sweden.

Key, M. R. (1975). *Male/female language.* New Jersey: The Scarecrow Press, Inc.

Lakoff, R. T. (1975). *Language and women's place.* New York: Harper & Row.

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

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.

Lenard, D. B. (2017). Gender differences in the personal pronouns use on the corpus of congressional speeches. *Journal of Research Design and Statistics in Linuistics and Communicaton Science, 3*(2), 161–188. Retrieved February 01, 2018, from https://bib.irb.hr/datoteka/861053.Gender_differences_in_the_personal_pronouns_use_of_the_corpus_of_congressional_speeches.pdf

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.

Mulac, A., Bradac, J. J., & Gibbons, P. (2001). Empirical support for the gender-as-culture hypothesis: An intercultural analysis of male/female language differences. *Human Communication Research, 27*(1), 121-152.

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

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: 10.1080/016385308020 73712

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

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:doi:10. 1037//0022-3514.77.6.1296

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*. Austin, TX: Pennebaker Conglomerates (www.LIWC.net).

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?AWSAcce

ssKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1509703685&Signature=up2ZF9XJMw22JMr%2FBxOYQXw40fg%3D&response-content disposition=inline%3B% 20filename% 3DSMS _Text_Analysis_Langu

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

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.language inindia.com/ june2010/v10i6june2010.pdf

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

Sinclair, J. (2005). 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/

Speer, S. A. (2005). *Gender talk: Feminism, discourse and conversation analysis*. New York: Routledge.

Spender, D. (1980). *Man made language.* London: Routledge.

Tausczik, Y. R., & Pennebake, 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

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.

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.

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

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

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/ BF0306 2342

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.