Tehseen Zahra[*]
Akhtar Abbas[**]

Development of Gender-based Pakistani Academic Word List (G-PAWL)

ABSTRACT

Differences in the language used by men and women have intrigued scholars since long. However, the use of corpus-based techniques for such research is infrequent, especially in Pakistan. After the advent of the concept of World Englishes, corpora are developed to show the identity of different countries, individuals, regions and areas of the world. Likewise, English spoken and written in Pakistan is different from the English language used in various countries/areas of the world. There is not only worldwide differences in spoken and written discourses, researches show that language used by men and women is also different. Therefore, this research aims at exploring lexical items used by men and women and ultimately developing Gender-based Pakistani Academic Word List (G-PAWL). The research articles from three disciplines of social sciences (Education, History and English) published in HEC recognized research journal of Y category were collected. Academic word lists, occurrences/frequencies of lexical items and contextual uses of lexicons by Pakistani writers in research papers are explored through corpus-based techniques. The academic word list of male writers comprises of more words (124 words) than female writers (37 words). Variation in contextual use of language is explored by using lexical item language. The findings demonstrate the variations in language used by men and women in Pakistani written academic discourse.

Keywords: Gender-based, Corpus, Academic Word List, Variation.

[*] Assistant Professor, Department of English, Air University, Islamabad, Pakistan

[**] Assistant Professor, Department of English, Air University, Islamabad, Pakistan

Introduction

Human language is not static and uniform but keeps on evolving with time. In consequence, language variation gradually develops due to several social variables. One of the most significant social variables is gender. Gender plays a significant role in generating language variation between men and women. A study carried out by Labov (1972) in New York also noted the differences and found that women in New York used less stigmatized forms in careful speech than men. Women are more conscious about the novel vocabulary and they were more sensitive to the use of eminent variants of languages. The study of language and gender has shifted the perspective of studying language variation to a heterogeneous diversity in style, manner, and use of lexical items. Corpus Linguists are interested in investigating where, when, and why differences in language emerge in discourse. Many researchers have employed corpus techniques to explore the variation in language used by men and women (Newman, Groom, Handelman, & Pennebaker, 2008; Baker, 2014).

Literature Review: Physical and Social Differences between Men and Women

There are two broad distinctions between men and women. One is the physical difference and the second is the social difference. Physically, females have more fat and fewer muscles and thus they are unable to do physical labour. On the other hand, males have more muscles and they are physically more fit so they are strong and they mature more rapidly and have a longer life. This traditional concept has been transferred over centuries that men are strong and thus they can work outside the comfortable atmosphere of the houses. This leads to the absolute authority of male over female and thus the logo "men outside, women inside" (Jinyu, 2014, p.2).

Socially, they have different skills and way of dealing with life. Females and males often show different advantages and skills in doing their work (Xia, 2013). There have been a lot of assumptions regarding male and female language in social spheres. However, the roots of this study were first planted by Lakoff groundbreaking work “language and

women place" in which he highlighted the differences in speaking between genders.

It would be, therefore, hypothetically incorrect to assume that men and women use the same language in discourses. Instead, even though the register is the same or different, the differentiation in gender brings a difference in language. One may notice that although hedges and other interactional discourse markers signal the same meanings, men and women perceive them differently. In consequence, we find the number of using these discourse markers different as well as the meanings they carry alongside them.

Language Used by Men and Women

Bell, Cole and Floge (1992) found that men prefer to use a more competitive style and women tend to exhibit a cooperative and facilitative style in their writings. Earlier, recommendation letters used to be written by men only. Men used to inculcate sexist elements towards women such as "highly decorative" or "personally attractive" (p.8). They carried out their corpus-based research on letters of recommendation within the domain of sociology and anthropology. The data revealed that men and women have different exposure to academics, and ultimately develop diverse use of linguistic items in letters of recommendation. Almost all the letters were written in a positive tone. This study compared how men and women write recommendation letters about the same candidate. The figure reveals the difference in the topics covered by male and female writers in letters of recommendations. The most recurring themes of female writers were teaching, research, collegiality, publications and administrative skills while male writers revealed the teaching skills, research potential, intellectual power and publications of the candidates respectively.

Topics Covered By Female and Male Writers in Letters of Recommendation, 1973–1984

Topic	Number of Writers Covering Topic	
	Women Writers (39)	Men Writers (39)
Teaching	36	35
Research	32	29
Intellect	14	30
Comparison	6	10
Intelligence	6	7
Quality	8	14
Intellectual Skills	2	14
Communication Skills	4	9
Collegiality	22	18
Publications	15	19
Personality	9	16
Administrative Skills	11	5
Career Potential	5	10

Figure 1: Topics covered by male and female writers in letters of recommendation, 1973-1984 (Source: Letters of recommendation in Academic: Do Women and Men Write in Different Languages by Susan E. Bell, C. Suzanne Cole and Liliane Floge)

Rubin and Kathryn Greene (1992) stated that difference in language is not due to the factor of gender difference alone, but is of the mode of discourse, too. A holistic aspect of discourse such as topic choices or modes of argumentation is significant. They built a corpus consisted of 88 candidates who were involved in voluntary writing sessions. In that session, 65 women and 32 men participated and they were asked to write argumentative as well as expressive writings. They found that female candidates (86 percent) used more exclamations than men (31 percent) in expressive writings. Some scholars consider exclamation marks a sign of 'excitability' that marks a departure from male writings. These exclamations signal intensification, such as "The spread of drugs on our campus has got to stop!" Such sentences, according to Rubin and Greene, convey a writer's lack of stature. Another significant discourse-level feature is acknowledging the legitimacy of opposing concerns such as "it's really important to reduce the number of students who are using drugs. Those are not only hurting themselves but others, too, still, mandatory drug testing is no solution" (Greene & Rubin). They found that more than half of women included such acknowledgement while only a quarter of men did. Flyn (1988) argues that women's writings are more affiliate than confrontational. Women used more egocentric sequences such as "I think", "I like", and "I felt" twice than men. However, men used more non-egocentric first person. The use of first person

pronoun in men's writings occurred in powerful expressions of assertion such as "I say we should do away with this idea of drug testing", "as a student of this university, I have a right to have my privacy protected" (p.29). Hence, one may notice that gender differentiation plays a pivotal role in constructing language. The gender differences describe certain attributes about males or females and the way language is used, modified, and manipulated. Moreover, it is sometimes observed that the way females assign meanings to discourse markers, perhaps never existed before, or are used differently by males in their writings.

Exploring Metadiscursive Patterns in Academic Discourse

Academic discourse is one of the most significant areas of research as it is closely intertwined with the practices of pedagogy and learning. It is therefore of relevant interest to probe into the variation in language used by males and females in academic discourse.

Following Hyland's taxonomy, Serhold (2012) has carried out a corpus-based metadiscourse analysis on the use of boosters and hedges in academic writing. The research aimed to explore the epistemic modality like hedges and boosters in written academic discourse based on gender differences. The material consisted of 20 essays; 10 female essays and 10 males. Each essay contained approximately 8000 or 9000 words. She found that the total number of occurrences of hedges in male writing was 442, and 403 in female's writings and a total number of 136 boosters were found more prevalent in female's writings than men's which contained 81 boosters. Hence, she concluded that men and women tended to use hedges and boosters in different manners that were not identical. For instance, men used the hedge 'suggest' to show the result of a phenomenon while women used 'suggest' to explain the hypothesis.

Salehi and Biria (2016) carried out an analysis of interactional metadiscourse markers that are used by men and women in introducing research articles that are taken from microbiology and applied linguistics, written by all-male, all-female, and co-authored by males and females. The reason for selecting these two disciplines is based on Becher and Trowler's idea of categorizing academic disciplines into soft and hard sciences. Since both of these disciplines are applied in nature, the researchers thus would pay great attention to them. A total number of 64 articles were selected, and each discipline carried out 32 articles that

consisted of 10 articles which were written by a male, 10 by female, and 12 were co-authored by both male and female. One-way ANOVA was used to explore the differences. Males and females used interactional markers differently with a percentage of 3.83 that is higher than the percentage of the interactional markers used by men with a percentage of 2.41. Although the study does not provide the readers with an example to illustrate the findings, the frequency of the discourse markers supports the above mentioned claim of language variation across genders.

Aragamon *et al* (2003) studied 604 samples from a large variety of genres in the British National Corpus and found that females used pronouns more than males who showed a greater tendency towards the use of noun specifiers e.g. determiners and numerals. Differences in the employment of pronouns and specifiers by the two sexes point to the fact that a relatively distinguishable strategy is being adopted by men and women in this regard. Pronouns are used when the author presumes that the audience is already familiar with the object that is being referred to. On the other hand, the employment of specifiers shows that the writer is supposing that the addressees are not completely aware of what he/she is referring to. Also, the outcomes of the above study suggest that females are probably more emotionally engaged in the discourse and are more attentive towards taking the charge of its course. Similarly, men exhibit more practicality by using figures and quantifiers and seem to be more interested in simply stating the state of affairs. This depicts that variation in the use of language by men and women exists in a non-negligible ratio and needs to be explored further at different levels of written discourse to make generalizations about the results.

	Fiction		**Nonfiction**	
Nominal type	**Female**	**Male**	**Female**	**Male**
Common nouns	1479	1596	2022	2061
Proper nouns	198	226	213	232
Pronouns	978	860	390	282
Total	**2655**	**2682**	**2625**	**2575**

Figure 2: Frequency of common nouns, proper nouns and pronouns in fiction and nonfiction (Adapted from Gender, genre, and writing style in formal written texts by S. Argamon, M. Koppel, J. Fine & A.R. Shimoni, 2003 *TEXT-THE HAGUE THEN AMSTERDAM THEN BERLIN*. P. 332)

Newman *et al* (2008, as cited in Ishikawa, 2015) gathered a huge corpus to examine gender differences in the employment of language. His data comprised of 3% speech samples and fiction from the seventeenth century. Their research proposes that females were found to be more in habit of using lexical items referring to the psyche and social norms and practices. Also, they exhibited a greater inclination towards the use of verbs. Men, on the other hand, were found to argue more about current affairs and used those lexical items that pointed towards material things and issues that were not private. The above study indicates that extensive use of pronouns by females exhibits their greater tendency to be indulged in social, psychological and familial practices. The results show that females use more pronouns, social words and psychological processes while men use more articles, numbers and prepositions.

Hyland & Tse (2008) conducted a corpus-based study of book reviews. They made a note of many variations between the language of male and female reviewers. Frequent uses of engagement markers, hedges, and self-mention were observed in the reviews of the male writers who had been writing the reviews of the books of female authors which was usually not a characteristic of the male style (Holmes, 1989). This was not the only difference found in the study. They also observed that 'transition markers' were the most frequent features in females' texts and the second most frequent elements in male's texts, 'hedges' and 'boosters' were frequently used by male interviewees and 'evidentials' and 'code glosses' were more frequent among female interviewees.

Academic Word List

Using academic vocabulary in academic discourses causes a great deal of difficulty for the researchers as well as learners as they are familiar with the technical vocabulary which is part of their course but have limited knowledge of academic lexical items. Coxhead (2000) considered this phenomenon and developed the Academic Word List (AWL). Coxhead built a corpus of 3.5 million words of research papers and university

textbooks covering four main areas, Natural Sciences, Law, Commerce and Arts. She applied the criteria of frequency (occur 100 times in the corpus), range (distribution of words in different disciplines) and specialized occurrence in the corpus (not included in General Service List (West, 1953)). Moreover, Coxhead (2000) divided AWL into different categories based on disciplines and word families

Wang, Liang and Ge (2008) also worked on the compilation of the Medical Academic Word List (MAWL). They collected data from online resources and developed the wordlist containing 623 word families. They found that some lexical items are used frequently in corpus and these words also covered wide text of medical research articles. Hence they concluded that these lexical items play a significant role in the communication of meaning. MAWL may also be useful in syllabus design especially for the compilation of course books of Medical academic vocabulary and for medical English learners who set their goals in learning particular vocabulary during English language learning.

Numbers of researches have been conducted on the compilation of academic word list. Mozaffari and Moini (2014) compiled an academic word list of education research articles. Lessard-Clouston (2013) built a corpus of research articles from different disciplines varying from agriculture, business, engineering, and medicine to theology. They developed an academic word list that was beneficial for EFL and ESL learners.

Review of the literature reveals the gender-based differences in using language in various discourses such as corpus-based meta discourse analysis on the use of boosters and hedges in essays (Serhold, 2012), uses of pronouns by men and women in British National Corpus (Aragamon et al, 2003) and uses of engagement markers, hedges, and self-mention in book reviews (Hyland & Tse, 2008). Furthermore, researches based on academic word lists are also revealed to elaborate the significance and need of academic word list. But, we found that limited research works have been conducted on the compilation of academic lexical items used by men and women in both written and spoken academic discourses. Hence, this research initially concentrates on the compilation of a gender-based corpus of research papers of three disciplines, Education, History and English and finally aims to develop a gender-based academic word list named Pakistani Academic Word List (PAWL).

Keeping in view the aims of the research paper, the following research questions are developed:

Q: What are the academic words used by men and women in Pakistani social sciences research papers?

Methodology

Gender variation and differences are found in almost every language. Studies identified that male and female language strategies are different which contain a systematic and functional variation. Many researchers have been conducted on a specific set of linguistic features used in academic writing (such as Serhold, 2012; Aragamon et al, 2003; Hyland & Tse, 2008). This study has been conducted in the Pakistani context to observe the linguistic variations of Pakistani male and female writers in academic discourse. The purpose was to propose Gender-based Pakistani Academic Word List (G-PAWL) and identify the linguistic variations exist between male and female in Pakistan.

The data for the current study consists of research papers written by Pakistani writers. This data is part of a larger research project named Pakistan Gender Text (PakGenText) (Shehzad and Zahra, 2019: in progress)**.** Weselected thirty research papers from three different disciplines (Education, History and English) of social sciences. The research papers were collected from electronic versions and published after 2017. The general format of the research papers was abstract, introduction, literature review, analysis and discussion. Fifteen of them were written by males and fifteen were written by females. The cleaned corpus of male and female writers comprised of 70,164 (Corpus M) words and 75,933(Corpus F) respectively. So, the total data comprised of nearly 146,097 words. For the standardization, the whole corpus was cleaned by removing tables, figures, footnotes, endnotes, bibliographies/references, acknowledgements and appendices. We selected research articles written by females and co-authored by females, males and co-authored by males. Research articles written/co-authored by males and females are not part of this study. The academic wordlist was explored through AntConc and content words carrying the frequency of 50 or more were added to the wordlist. Although researchers used articles, prepositions, pronouns, nouns and helping verbs, this study was restricted to content

words only. The frequencies of lexical items are given in the tables of wordlists in brackets after every lexical item (See Appendices).

Analysis

Initially, the academic word list is explored in the whole corpus of male and female writers. Further, lexical items are explored in both corpora of male and female writers separately to compile academic wordlist in each corpus. Frequently occurring words carrying the frequency of more than fifty are explored at three levels.

1. Collective Academic wordlist (Attached as Appendix A)
2. Academic wordlist of female writers (Attached as Appendix B)
3. Academic wordlist of male writers (Attached as Appendix C)

We found a total of thirty seven content words that occur fifty times or more in the corpus of female writers. These frequencies of occurrences reveal those female writers have variation in language use among them as well. First ten frequently occurring words in corpus F are Children (242), Language (241), Hearing (131), District (119), Teachers (117), Education (114), Parents (106), Study (106), Learning (100) and School (100). The frequencies of the words also reveal the prevalent themes in Corpus F. The ubiquitous theme in female research papers are related to the languages used by children, the role of teachers and parents in Children's education, the role of hearing in learning and school education in various districts.

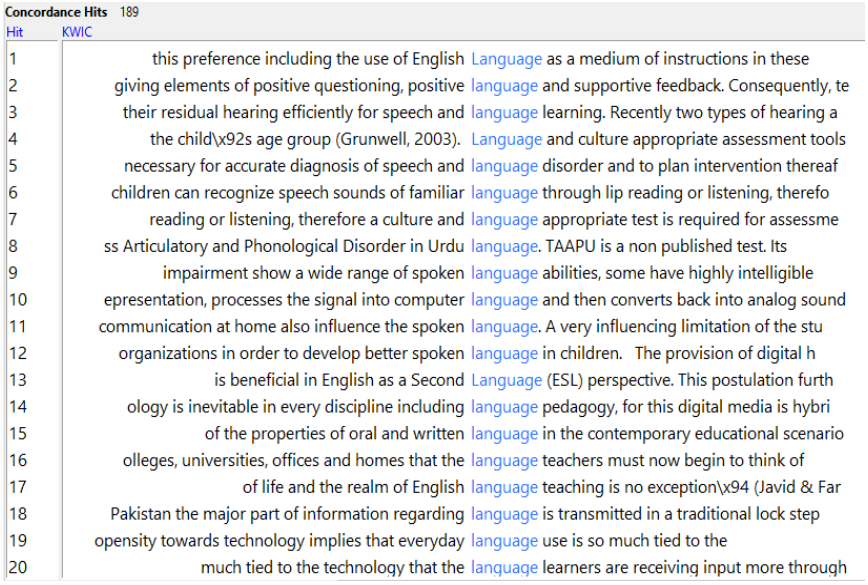
On the other hand, first ten frequently occurring words in Corpus M are Translation (275), Teachers (244), Language (236), English (218), Pakistani (193), Study (193), Cultural (158), Text (153), Such (148) and Students (147). Hence the prevalent themes are the role of translation in English language teaching in Pakistan, Cultural aspects of language and representation of Pakistani Culture through text.

The word lists of male and female writers clearly show the variation in language use. In both corpora (Corpus M and Corpus F), we selected the content words occurring 50 times in each corpus. A significant difference is observed as far as the frequency of the content words is concerned like the wordlist of male writers consists of 124 words while the wordlist of female writers consists of 37 content words. This difference advocates the idea that female writers use limited lexical items to express their ideas in various contexts or they might have limited vocabulary to express themselves in

academic discourse. While the wordlist of male writers consists of 124 words which is more than three times greater than male writers. Hence, it is assumed that male writers, on one hand, use diverse lexical items as compared to female writers and on the other hand, they may have discussed diverse themes in research papers. The lexical items carrying the frequency of 50 or more are higher in number in Corpus M than Corpus F that shows high similarity or repetition of words. Therefore, it is assumed that male writers repeat lexical items frequently than female writers. Later we merged both the corpora of male and female writers to get an overall academic wordlist. Overall, 235 words occur fifty times in this corpus (See Appendix A). The wordlist shows the lexical items used by men and woman in general and also the prevalence of various themes in three disciplines (Education, History and English). Interestingly, we also noticed the existence of some lexical items in both corpora (Corpus M and Corus F): these lexical items are *language, teachers, education, study, learning, school, process, more, between, only, some, level, public, both, development, analysis, research, results* and *different*.

The study of right and left collocates also show the variation in language use between men and women. Further, to testify this stance, we browsed lexical item language in both corpora (Corpus F and Corpus M). We selected a node word *language* as it is a frequently occurring word in academic word list A and it also exists in both corpora i.e. Corpus M and Corpus F. We browsed *language* in AntConc using corpus F and explored the themes fore grounded in the first twenty hits. The prevalent themes are language as a medium of instruction (occur in hit no 1, 2 and 17), evaluative use of language (occur in hit no 2 and 4), language learning (occur in hit no 3 and 20), speech and language disorder (occur in hit no 5, 8 and 9), speech recognition (occur in hit no 6), computational aspects of language (hit no 10 and 19), use of language in home (hit no 11) and language pedagogy (hit no 11, 15, 16 and 18).

Figure 3: First twenty occurrences of *language* in the corpus of female writers (Corpus F)

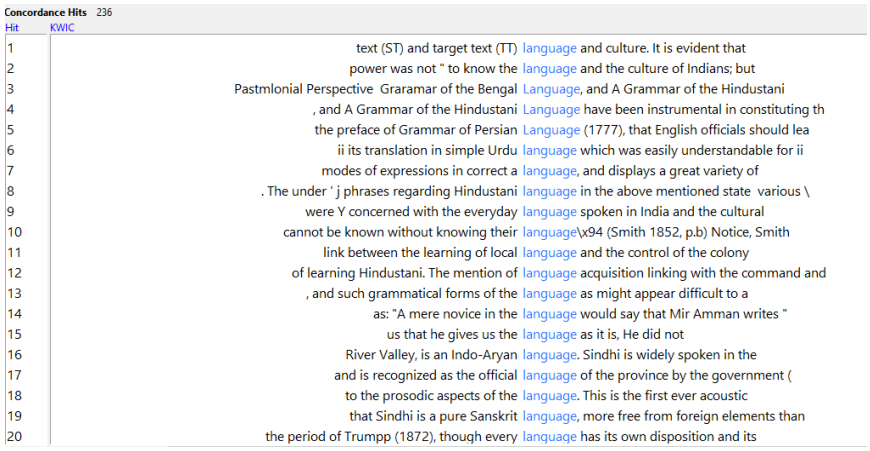


Concordance Hits 189

Hit	KWIC		
1	this preference including the use of English	Language	as a medium of instructions in these
2	giving elements of positive questioning, positive	language	and supportive feedback. Consequently, te
3	their residual hearing efficiently for speech and	language	learning. Recently two types of hearing a
4	the child\x92s age group (Grunwell, 2003).	Language	and culture appropriate assessment tools
5	necessary for accurate diagnosis of speech and	language	disorder and to plan intervention thereaf
6	children can recognize speech sounds of familiar	language	through lip reading or listening, therefo
7	reading or listening, therefore a culture and	language	appropriate test is required for assessme
8	ss Articulatory and Phonological Disorder in Urdu	language	. TAAPU is a non published test. Its
9	impairment show a wide range of spoken	language	abilities, some have highly intelligible
10	epresentation, processes the signal into computer	language	and then converts back into analog sound
11	communication at home also influence the spoken	language	. A very influencing limitation of the stu
12	organizations in order to develop better spoken	language	in children. The provision of digital h
13	is beneficial in English as a Second	Language	(ESL) perspective. This postulation furth
14	ology is inevitable in every discipline including	language	pedagogy, for this digital media is hybri
15	of the properties of oral and written	language	in the contemporary educational scenario
16	olleges, universities, offices and homes that the	language	teachers must now begin to think of
17	of life and the realm of English	language	teaching is no exception\x94 (Javid & Far
18	Pakistan the major part of information regarding	language	is transmitted in a traditional lock step
19	opensity towards technology implies that everyday	language	use is so much tied to the
20	much tied to the technology that the	language	learners are receiving input more through

We browsed a node word language through AntConc in Corpus M and explored the themes foregrounded in the first twenty hits. The prevalent themes are language and culture (occur in hit no 1, 2, 9, 11, 13 and 16), grammatical aspects of language (occur in hit no 3, 4, 5 and 8), language and translation (occur in hit no 6), language errors (occur in hit no 7), language learning and acquisition (occur in hit no 12), language use (occur in hit no 14, 15 and 17) and features of language (occur in hit no 18, 19 and 20).

Figure 4: First twenty occurrences of *language* in the corpus of male writers (Corpus M)



Concordance Hits 236

Hit	KWIC		
1		text (ST) and target text (TT)	language and culture. It is evident that
2		power was not " to know the	language and the culture of Indians; but
3		Pastmlonial Perspective Graramar of the Bengal	Language, and A Grammar of the Hindustani
4		, and A Grammar of the Hindustani	Language have been instrumental in constituting th
5		the preface of Grammar of Persian	Language (1777), that English officials should lea
6		ii its translation in simple Urdu	language which was easily understandable for ii
7		modes of expressions in correct a	language, and displays a great variety of
8		. The under ' j phrases regarding Hindustani	language in the above mentioned state various \
9		were Y concerned with the everyday	language spoken in India and the cultural
10		cannot be known without knowing their	language\x94 (Smith 1852, p.b) Notice, Smith
11		link between the learning of local	language and the control of the colony
12		of learning Hindustani. The mention of	language acquisition linking with the command and
13		, and such grammatical forms of the	language as might appear difficult to a
14		as: "A mere novice in the	language would say that Mir Amman writes "
15		us that he gives us the	language as it is, He did not
16		River Valley, is an Indo-Aryan	language. Sindhi is widely spoken in the
17		and is recognized as the official	language of the province by the government (
18		to the prosodic aspects of the	language. This is the first ever acoustic
19		that Sindhi is a pure Sanskrit	language, more free from foreign elements than
20		the period of Trumpp (1872), though every	language has its own disposition and its

The study of right and left collocates of *language* reveal the variations in language, as well as themes prevalent in Corpus M and Corpus F. For example in the corpus of female writers, different collocates, are used with *language* such as *the use of English language as a medium of instruction, positive language and supportive feedback, speech and language learning, language and culture, speech and language disorder, the sound of familiar language through lip reading*. Moreover, different collocates are used with *language* in the corpus of male writers such as *target text language and culture, know the language and culture of Indians, Grammar of Bengal language, the grammar of Hindustani language have been instrumental, correct a language, everyday language spoken in India, learning of local language*. Variations in using the right and left collocate with a node word *language* not only support the idea that men and women use different lexica items but also reveal the variations in themes of research articles written by men and women.

Recommendations

Limited researches have been conducted on Gender-based use of language in Pakistani discourses using corpus related methodologies so far. Hence, there is a dire need to explore Pakistani discourses using corpus linguistics. We selected thirty research papers from three different disciplines, Education, History and English for this research. Data can be collected from some other disciplines (such as Chemistry, Physics, Islamic Studies) to make word lists of these disciplines. Moreover, this study

focuses on consolidated word lists based on three disciplines. Separate wordlists of different disciplines may also be developed by building sub-corpora of each discipline. Moreover, in this study, we focus on a single lexical item; *language* to explore various themes that fore grounded this word. Some more researches may be conducted by analyzing gender-based uses of language like parts of speech, thematic evaluation of language and contextual use of language. Data may be collected from various resources covering Pakistani discourses to explore gender-based uses of language such as blogs, newspapers, books, theses (of various disciplines) and letters to editors. Moreover, data can be collected from spoken discourses such as morning shows, classroom lectures, conference presentations, seminars, motivational speeches to compile a word list and explore contextual use of language.

Conclusion

The results show variations in lexical items used by men and women in written academic discourse. The academic word list of male writers comprises of more words (124 words) than female writers (37 words). The high frequencies of some lexical items used by male writers show that men tend to repeat the same lexical items frequently. Moreover, the numbers of words in academic word lists of male and female writers reveal that men tend to use new vocabulary to express various themes. Similarly, there is variation in the contextual use of language. It is explained with a simple example of *language*, where women use it in various themes such as language as a medium of instruction, evaluative use of language, language learning, speech and language disorder, speech recognition, computational aspects of language, use of language in home and language pedagogy while men use it (*language*) in a different context such as language and culture, grammatical aspects of language, language and translation, language errors, language learning and acquisition, language use and features of the language. So, the study of collocates also reveals the dominant themes discussed by males and females in their research papers. In a nutshell, it is concluded that heterogeneity is a special characteristic in gendered language. The use of language mirrors gender differences in style and also gender difference reflects how language is used, modified, codified, and manipulated, thus, bringing variation in language with a great diversity in meanings and functions of language.

References

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. Retrieved from https://doi.org/10.1515/text.2003.014

Baker, P. (2014). *Using corpora to analyze gender*. London: Bloomsbury Academic.

Bamman, D., Eisenstein, J., & Schnoebelen, T. (2014). Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, *18*(2), 135-160. Retrieved from https://doi.org/10.1111/josl.12080

Bell, S. E., Cole, C. S., & Floge, L. (1992). Letters of recommendation in academe: Do women and men write in different languages? *The American Sociologist*, *23*(3), 7-22. Retrieved from https://doi.org/10.1007/BF02691910

Biber, D., & Conrad, S. (2001). Quantitative corpus-based research: Much more than bean counting. *TESOL Quarterly*, *35*(2), 331-336. Retrieved from https://www.jstor.org/stable/3587653

Coxhead, A. (2000). A new academic word list. *TESOL quarterly*, *34*(2), 213-238.

Ghafoori, N., & Oghbatalab, R. (2012). A comparative study of Metadiscourse in academic writing: Male vs. female authors of research articles in applied linguistics. *Journal of English Language Pedagogy and Practice*, 5(10), 87-113. Retrieved from http://jal.iaut.ac.ir/article_520094.html

Holmes, J. (1990). Hedges and boosters in women's and men's speech. *Language & Communication*, 10(3), 185-205. Retrieved from http://www.sciencedirect.com/science/article/pii/027153099090002S

Holmes, J. (1995). Women, men, and politeness. London: Longman.

Holmes, J. (2001). An introduction to Sociolinguistics (2nd ed.). London: Longman.

Hyland, K. (2005). Metadiscourse: exploring interaction in writing. London: Continuum.

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173-192. Retrieved from https://doi.org/10.1177/1461445605050365

Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics*, *25*(2), 156–177. Retrieved from https://doi.org/10.1093/applin/25.2.156

Jinyu, D. (2014). Study on gender differences in language under the sociolinguistics. *Canadian Social Science*, *10*(3), 92-96.

Labov, W. (2011). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (n.d.). Transmission and diffusion. *Language*, *83*(2), 7-23.

Lakoff, R. (1973). Language and woman's place. *Language in Society* , *2*(1), 45-80. Retrieved from https://www.jstor.org/stable/4166707

Lessard-Clouston, M. (2013). Word Lists for vocabulary learning and teaching. *CATESOL Journal*, *24*(1), 287-304.

Mauranen, A. (2002). 'A good question': Expressing evaluation in academic speech. Domain-specific English: Textual practices across communities and classrooms, 115-140.

Mozaffari, A., & Moini, R. (2014). Academic words in education research articles: A corpus study. *Procedia-Social and Behavioral Sciences*, Conference Proceedings, *98*, 1290-1296.

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. doi:10.1080/01638530802073712

Rubin, D. (1992). Gender-typical style in written language. *Research in the Teaching of English*, *26*(1), 7-40. Retrieved from http://www.jstor.org/stable/40171293

Salehi, B. M., & Biria, R. (2016). Exploring gender differences in cross-disciplinary discourse: Interactional Metadiscourse markers in the discussion section of research articles. *Language in India*, *16*(12), 169-186. Retrieved from

http://www.languageinindia.com/dec2016/salehimetadiscourse1.html

Schleef, E. (2008). Gender and academic discourse: Global restrictions and local possibilities. *Language in Society*, *37*(4), 515-538. Retrieved from https://www.jstor.org/stable/20108177

Serholt, S. (2012). Hedges and boosters in academic writing - A study of gender differences in essays written by Swedish advanced learners of English. Retrieved from http://hdl.handle.net/2077/29526 (Student Essay)

Shehzad,W and Zahra, T (2019, Ongoing research project). Pakistan Gender Text (PakGenText). Funded by Higher Education Commission Pakistan.

Tannen, D. (1996). *Gender and discourse*. New York: Oxford University Press.

Tannen, D. (2013). *You just don't understand: Women and men in conversation*. London: Virago.

Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society*, *1*(2), 179-195. Retrieved from http://www.jstor.org/stable/4166683

Wang, J., Liang, S. L., & Ge, G. C. (2008). Establishment of a medical academic word list. *English for Specific Purposes*, *27*(4), 442-458.

Xia, X. (2013). Gender differences in using language. *Theory and Practice in Language Studies*, *3*(8), 1485-1489. doi:10.4304/tpls.3.8.1485-1489

Xiao, R., & Tao, H. (2007). A corpus-based sociolinguistic study of amplifiers in British English. *Sociolinguistic Studies*, *1*(2), 241-273. doi:10.1558/sols.v1i2.241

Yeganeh, M. T., & Ghoreyshi, S. M. (2015). Exploring gender differences in the use of discourse markers in Iranian academic research articles. *Procedia - Social and Behavioral Sciences*, *192*, 684-689. Retrieved from https://doi.org/10.1016/j.sbspro.2015.06.104

Appendix A: Overall Academic Wordlist

1.	Language	425	78.	way	101	155.	Political	65
2.	Teachers	361	79.	what	101	156.	According	64
3.	Translation	358	80.	because	100	157.	Basis	64
4.	Study	299	81.	government	99	158.	Following	64
5.	All	275	82.	state	99	159.	Good	64
6.	But	268	83.	if	98	160.	Life	64
7.	English	265	84.	based	97	161.	Practices	64
8.	Such	255	85.	formal	96	162.	Shows	64
9.	Children	251	86.	materials	96	163.	Society	64
10.	Other	239	87.	processes	95	164.	Terms	64
11.	School	236	88.	like	94	165.	Various	64
12.	One	228	89.	therefore	94	166.	Building	64
13.	Schools	227	90.	any	93	167.	Child	63
14.	There	224	91.	educational	93	168.	Practice	63
15.	Education	216	92.	words	92	169.	Skills	63
16.	Pakistani	206	93.	class	91	170.	View	63
17.	More	202	94.	table	91	171.	Fact	62
18.	Students	190	95.	call	89	172.	Means	62
19.	Pakistan	189	96.	nature	89	173.	Physical	62
20.	Cultural	186	97.	respondents	89	174.	Situation	62
21.	Only	186	98.	business	87	175.	Stress	62
22.	Text	184	99.	difference	87	176.	Digital	61
23.	Time	183	100.	features	87	177.	Given	61
24.	Learners	182	101.	first	87	178.	Information	61
25.	Between	178	102.	informal	87	179.	Local	61
26.	Both	170	103.	present	87	180.	Satisfaction	61
27.	Found	167	104.	motivation	86	181.	Approach	60
28.	Some	164	105.	European	85	182.	Do	60
29.	Different	163	106.	Factors	85	183.	Further	60
30.	Analysis	160	107.	New	85	184.	Politics	60
31.	Role	158	108.	Using	85	185.	Categories	59
32.	Learning	157	109.	Performance	84	186.	Native	59
33.	Teacher	157	110.	Power	84	187.	Purpose	59
34.	Process	155	111.	Speech	84	188.	Value	59
35.	Social	150	112.	Positive	83	189.	Another	58
36.	Use	149	113.	Could	82	190.	Case	58
37.	Teaching	148	114.	Community	81	191.	Items	58
38.	While	147	115.	Related	81	192.	Participants	58
39.	Data	143	116.	Candidate	79	193.	During	57
40.	Than	141	117.	After	78	194.	Syllable	57

41. Level 140	118. Majority 78	195. Analog 56
42. People 139	119. National 78	196. Experience 56
43. Used 139	120. Practicum 78	197. Part 56
44. Research 138	121. System 77	198. Understanding 56
45. Reportage 137	122. Texts 77	199. Computer 55
46. District 135	123. Made 76	200. Law 55
47. Linguistic 135	124. Translator 75	201. Point 55
48. Two 135	125. Members 74	202. Primary 55
49. Hearing 133	126. Pitch 74	203. Thus 55
50. Activities 132	127. Sector 74	204. Does 54
51. Press 132	128. Head 73	205. Election 54
52. Most 130	129. Towards 73	206. Focus 54
53. Well 130	130. Under 73	207. Four 54
54. Who 129	131. Religious 72	208. Retention 54
55. Among 125	132. Same 72	209. Support 54
56. faculty 125	133. British 71	210. Take 54
57. development 125	134. Due 71	211. Those 54
58. so 123	135. How 71	212. Especially 53
59. should 122	136. Intelligence 71	213. Example 53
60. source 122	137. Languages 71	214. Groups 53
61. however 118	138. Needs 71	215. Including 53
62. urdu 118	139. Colonial 70	216. Make 53
63. public 117	140. Emotional 70	217. Making 53
64. when 116	141. Significant 70	218. Mean 53
65. media 115	142. Target 70	219. Individual 53
66. results 114	143. Aids 69	220. Policy 52
67. work 112	144. Leadership 69	221. Sample 52
68. important 111	145. Art 68	222. Second 52
69. literature 111	146. Cases 68	223. Better 51
70. authentic 110	147. Translators 68	224. Courseware 51
71. culture 109	148. Word 68	225. High 51
72. need 107	149. Form 67	226. Translated 51
73. many 106	150. Context 66	227. Upon 51
74. studies 104	151. Institutions 66	228. Working 51
75. register 103	152. Major 66	229. Against 50
76. world 103	153. Years 66	230. Assessment 50
77. help 102	154. Articulation 65	231. Available 50
78. texts 77		232. Basic 50
		233. Design 50
		234. Differences 50
		235. Facilities 50

Appendix B: Word List of Female Writers

1. Children (242)	21. Emotional (64)
2. Language (189)	22. Child (63)
3. Hearing (131)	23. Level (63)
4. District (119)	24. Public (62)
5. Teachers (117)	25. Aids (60)
6. Education (114)	26. Both (60)
7. Parents (106)	27. Development (60)
8. Study (106)	28. Digital (60)
9. Learning (100)	29. Analysis (58)
10. School (100)	30. Analog (56)
11. Process (77)	31. Research (56)
12. State (74)	32. Educational (55)
13. Speech (73)	33. Results (54)
14. Intelligence (71)	34. Building (52)
15. More (70)	35. Different (52)
16. Need (70)	36. Courseware (51)
17. Between (67)	37. Difference (51)
18. Only (67)	
19. Some (66)	
20. Articulation (65)	

Appendix C: The Word List of Male Writers

1.	Translation	275	42.	time	91	83.	literature	67
2.	Teachers	244	43.	while	89	84.	because	65
3.	Language	236	44.	people	88	85.	translators	65
4.	English	218	45.	most	87	86.	development	64
5.	Pakistani	193	46.	important	85	87.	performance	63
6.	Study	193	47.	who	85	88.	translator	63
7.	Cultural	158	48.	business	84	89.	when	63
8.	Text	153	49.	way	84	90.	art	62
9.	Such	148	50.	European	83	91.	factors	61
10.	Students	147	51.	however	83	92.	stress	61
11.	School	140	52.	research	82	93.	many	60
12.	Reportage	137	53.	world	82	94.	results	60
13.	More	132	54.	respondents	81	95.	after	59
14.	Press	132	55.	features	80	96.	first	59
15.	About	130	56.	candidate	79	97.	politics	58
16.	Through	128	57.	nature	79	98.	learning	57
17.	Schools	127	58.	no	79	99.	own	57
18.	Found	121	59.	well	79	100.	related	57
19.	Linguistic	120	60.	practicum	78	101.	where	57
20.	Teaching	120	61.	process	78	102.	above	55
21.	Only	119	62.	level	77	103.	public	55

22. Faculty 118	63. culture 76	104. activities 54
23. Between 111	64. like 76	105. help 54
24. Different 111	65. power 76	106. leadership 54
25. Use 111	66. pitch 73	107. participants 54
26. both 110	67. very 73	108. what 54
27. social 110	68. class 72	109. even 53
28. source 110	69. words 72	110. given 53
29. authentic 107	70.among 71	111. members 53
30. Pakistan 105	71. motivation 71	112. native 53
31. role 105	72. out 71	113. religious 53
32. media 103	73. any 70	114. retention 53
33. analysis 102	74. British 70	115. situation 53
34. education 102	75. head 69	116. categories 52
35. register 102	76. new 69	117. means 52
36. used 100	77. studies 69	118. over 52
37. data 98	78. therefore 69	119. community 51
38. some 98	79. colonial 68	120. due 51
39. learners 96	80. texts 68	121. if 51
40. materials 95	81. work 68	122. election 50
41. two 92	82. government 67	123. made 50
		124. major 50