



Language and Gender in Pakistani Newspaper: Use of Interactional Metadiscourse Markers in Male and Female Bloggers

Adeel Khalid	Lecturer, Forman Christian College (A Chartered University)/Ph.D. Scholar Islamic International University, Islamabad, Pakistan.
Abrar Ajmal	Assistant Professor, English Language Center, Minhaj University, Lahore, Punjab, Pakistan.
Aqsa Javed	Lecturer, English Language Center, Minhaj University, Lahore, Punjab, Pakistan.

Vol. 2, No. 3 (Summer 2022)

Pages: 119 – 129

ISSN (Print): 2789-441X

ISSN (Online): 2789-4428

Key Words

Gendered Corpora, Interactional Metadiscourse Markers, AntConc Software, Pakistani Newspaper, Language Analysis

Corresponding Author:

Abrar Ajmal
Email: abrar.ajmal@mul.edu.pk

Abstract: *In newspapers, opinionated blogs written by male and female writers reflect subtle gender differences. This research aimed to study these gender differences through the lens of Hyland's (2005) framework of interactional metadiscourse markers used by male and female bloggers in their writings. For this study, two medium-sized gendered corpora have been developed containing 1,23,847 words, and the data was retrieved from the Dawn newspaper from January 2020 to December 2020. To analyze this data interactional metadiscourse markers were identified, and the frequency pattern of such instances had been noted in the blogs through corpus analysis. The free software tool AntConc was used to evaluate the frequencies of the selected expressions in male and female corpora. The findings indicated that female bloggers were more inclined towards using hedges, personal pronouns, boosters interactional metadiscourse markers in their language as compared to male bloggers.*

Introduction

In newspapers, opinionated blogs written by male and female writers reflect subtle gender differences. Such blogs have been harnessed to extract information to study gender differences in different contexts. Such data can be exploited conveniently to see how written discourse reflects gender differences in a text. These differences can be of various types ranging from content to the function words and this invokes the researchers to investigate what those differences are and how men and women tend to position differently as referred by Deborah Tannen's approach to studying gender difference (Newman, Groom, Handleyman and Pennybaker, 2008).

This paper attempts to study these gender differences by using Hyland's model (2005) of interactional metadiscourse markers (MDMs) by building two medium-sized corpora, one for male bloggers and the other for female bloggers (Hyland, 2005). The term originally used by Harris (1970), is an umbrella term that refers to the simple as "writing about writing, whatever does not refer to the subject matter being addressed". He claims that metadiscourse features provide a way of talking to the reader about the subject matter or propositional content. Metadiscourse has frequently been categorized in previous research about the three communicative language functions described by Hallidayan

systemic functional grammar. (e.g., Halliday, 1994), as has been pointed out (e.g., Ädel, 2006, Hyland, 2005). Hyland explains what they are. (2005, p. 26) in the following terms:

The Ideational Function: The representation of experience and ideas through language.

The Interpersonal Function: the process of encoding interaction through language that enables us to communicate with others, assume roles and express and comprehend sentiments and judgments

The Textual Function: the process of using language to structure a document so that it makes sense to both readers and the rest of the world.

Although some research studies have been conducted by various researchers worldwide within the Pakistani context, there is a need to research this domain for the use of MDMs to study gender differences. This study aims to bridge this gap in the existing literature while studying the language of male and female bloggers in Pakistani newspapers.

Statement of the Problem

According to the social constructionist theory, gender is a social and cultural construction that "is done every time we talk" (Coates, 2004, p.7). Men and women may be distinguished from one another only by innate biological and psychological characteristics that are contested by this "fluid" approach to language and gender (Speer, 2005, p. 13). This binary gender divide is the root of the Deborah Tannen-credited difference approach. This study lays its basis on the same idea of difference approach that supports the claim that the language used by men and women differ significantly. It necessitates additional research to ascertain if the language of these gendered blogs supports the difference approach or not.

Research Objectives and Research Questions

The research objective is to highlight the difference in language used concerning the usage

of interactional metadiscourse markers (MDMs) by the male and female Op-Eds authors in their written discourse of Pakistani Newspapers and to address this objective, this study aims at answering the following question:

- What is the difference in male and female Op-Eds' use of interactional metadiscourse markers (MDMs) in their written language of Pakistani Newspapers?

Literature Review

Lakoff's seminal work in studying gender differences laid the foundation in which she suggests that powerless speech uses more tag questions, and hedges, and intensifies when it is compared to powerful speech (Lakoff, 1973). She refers to women's language as less prestigious than that of men. Tannen studies gender and language and presents her different approach in her work *You Just Don't Understand: Women and Men in Conversation* (1990). Her approach helps in developing two separate cultural models of male and female in which children are allowed to socialize based on their gender (Tannen, 1990, p.47). Talbot argues that differences in language use between men and women are genetically determined and therefore natural (Talbot, 2010).

In the field of corpus linguistics, gender differences have been studied by various researchers and academicians. Newman, Groom, Handleyman, and Pennybaker (Newman et al., 2008) analyze gender differences using the standardized category to study a record of 14,000 text files from different studies and conclude that the female gender is inclined towards following psychological and social procedures whereas the male gender is inclined towards following object properties and impersonal topics. Biber and Burges (2000) explore the differences in linguistic behavior patterns of males and females in the works of the past 20 years and deduct distinct male and female features from plays written by the two genders from the developed corpus. (Biber & Burges, 2000). Schmid researches BNC

spoken text to find out gender differences and finds that females tend to use color words, temporal adverbs, and hedges more than males (Schmid, 2000). Baker (2014) investigates language and gender in the field of corpus linguistics by using the case study technique and finds in the patterns of directives in spoken conversations in sexist and non-sexist language use and how the press represents gay men, and the ways that male and female genders are constructed through language (Bloomsbury.com, n.d.). Khan and Afsar (2019) analyze gender differences and the language used while developing newspaper corpora and note certain subtle gender differences in the language used by the bloggers (Khan & Afsar, 2019).

This study is an attempt to analyze gender differences in the newspaper discourse within the field of digital journalism in Pakistan which is an emerging domain in the field of digital humanities and applied and corpus linguistics. This research hopes to fill the gap in the field of corpus linguistics by providing methodological underpinnings and informing the researchers and policymakers to be mindful of the fact that gender is a sensitive issue and should be dealt with appropriately. It is also to raise awareness of gender differences among the masses. Consequently, this research will find an answer to the following research question: How frequently do female and male bloggers use MDMs in their blogs in the leading newspaper of Pakistan?

Research Methodology

Theoretical Framework

This study is based on the model of Hyland's MDMs (Hyland, 2000) that he classifies into two sub-categories i.e., 'interactional' and 'interactive' and the scope of the study focuses only on the former category. The interactional category considers the interactional dimension of metadiscourse, which is concerned with "the readers' involvement in the text (Hyland, 2004)" and "the writer's efforts to control the level of personality in a text and establish a suitable

relationship to his or her data, arguments, and audience entails the following five groups of markers. Hyland identifies these interactional discourse markers into five types namely hedges, attitude markers, boosters, and use of pronouns (Hyland, 2005). The scope of this research and to address the posed research question, this study will focus on this interactional dimension of MDMs.

Corpus Building Procedure and Tool

The data collection procedure involved the creation of these two medium-sized corpora that had been used for this study. These electronic texts allowed their readers a rich source of data to explore linguistic differences used in male and female bloggers. Corpora for such studies were specialized and were developed to meet the specific needs of the research objectives (Androutsopoulos & Beißwenger, 2008). Hence, two specialized corpora were developed especially for the present study to compare the language of male and female writers. The specific issues of representativeness and size were considered while developing these two gendered corpora.

For this study, two male and two female blog writers had been chosen based on their age, and nationality available in the blogs of the E-newspaper of the leading newspaper in Pakistan i.e., Dawn. The researcher collated and included all the blog post entries of selected male and female writers from the previous year starting from January 2020 till December 2020. The reason for the selection of Dawn newspaper was that it is one of the reputed on national and international fronts and the renowned intelligentsia of Pakistan regularly contribute their opinionated editorials generally known as OP-EDs for the newspaper. One reason for its reputation is that it was founded by Quaid-e-Azam Muhammad Ali Jinnah.

These blog posts were stored from the archives in Microsoft Word file format from the website; was done manually by the researcher. Each blog post was stored and sorted out date-wise as a separate MS Word file under a number

and arranged in date showing the blogger's name initial letters. All the Sorting out of the data was done manually to make it representative of the target population. Two separate folders were maintained: one for female corpus and the other for male corpus. The data had been retrieved from January 2020 to December 2020. After sorting it out cleaning the data to remove noise was the second phase (Subramaniam et al., 2009) to make the text free from any ads, internet links, or hidden hyperlinks. The next step was to convert all these blog entries from MS Word into a Text Document file before proceeding to the analysis phase.

For building such specialized corpora, bloggers' profiles helped in addressing the issue of representativeness of the target population e.g., nationality, age, and gender to qualify the criteria determined by the external factors i.e., determine the population boundaries and hierarchal organization within the population (Biber, 1993). For this study, the language of the blog was English, blog type was Pakistani newspaper opinionated editorials written by four Pakistani adult bloggers two males and two females. As far as the size of this corpus is

concerned a total of 128 posts were included out of which 64 were female bloggers containing 58544 and 64 were male blog writers containing 64,341 words. So, in total, these corpora contained 1,22,885 words. The corpus analysis was conducted by using the corpus tool AntConc (Anthony, 2017) which is freely available and developed by Laurence Anthony. Ethical considerations have been followed throughout carrying out this academic project as the identity of the blog writers has been kept confidential throughout. The source of data has been acknowledged.

Results and Findings

For the research question of this study, a manual corpus analysis has been run to locate the frequency of each MDM available in these two gendered corpora. The data has been analyzed by the researchers of this study to avoid any mistakes in detecting and calculating the number and type of MDMs in the whole corpus. Then the frequencies of both corpora have been compared to examine the use of these interactional discourse markers in the language of male and female bloggers.

Table 1. Word Tokens and Word Types

	Male-Corpus	Female-Corpus
Word Tokens	65,353	58494
Word Types	6741	7705

Table 2. Use of Interactional MDMs

Use of Pronouns	M-Corpus	% Age	F-Corpus	% Age
First-Person Pronouns				
I	99	0.15%	173	0.30%
Me	14	0.02%	51	0.09%
Mine	1	0.00%	1	0.00%
My	21	0.03%	65	0.11%
Our + Ours	67	0.10%	112	0.19%
Ourselves	3	0.00%	9	0.02%
Us	91	0.14%	98	0.17%
We	232	0.35%	158	0.27%
Total	528	0.81%	667	1.14%

Second Person Pronouns				
You	110	0.17%	39	0.07%
Your	15	0.02%	8	0.01%
Total	125	0.19%	47	0.08%
Third Person Pronouns				
He	151	0.23%	141	0.24%
She	7	0.01%	118	0.20%
Their	311	0.48%	229	0.39%
Themselves	14	0.02%	35	0.06%
They	324	0.50%	323	0.55%
Total	807	1.23%	846	1.45%
Use of Hedges				
About	84	0.13%	99	0.17%
Almost	24	0.04%	7	0.01%
Apparent + Apparently	6	0.01%	14	0.02%
Appropriate + Appropriately	2	0.00%	4	0.01%
Argue + Argued + Argues	31	0.05%	4	0.01%
Around	73	0.11%	26	0.04%
Assume + Assumed + Assumes	3	0.00%	3	0.01%
Believe + Believed + Believes	11	0.02%	33	0.06%
Broadly	2	0.00%	0	0.00%
Claim + Claimed + Claims	8	0.01%	11	0.02%
Could	56	0.09%	101	0.17%
Couldn't + Could not	0	0.00%	7	0.01%
Doubt	8	0.01%	5	0.01%
Essentially	0	0.00%	11	0.02%
Estimate + Estimated + Estimates	14	0.02%	3	0.01%
Fair + Fairly	9	0.01%	2	0.00%
Feel + Feels + Felt	13	0.02%	25	0.04%
Frequently	2	0.00%	0	0.00%
Generally	6	0.01%	8	0.01%
Guess	2	0.00%	2	0.00%
In general	1	0.00%	6	0.01%
In my opinion	1	0.00%	0	0.00%
In my view	0	0.00%	1	0.00%
Indicate + Indicated + Indicates	4	0.01%	2	0.00%
Largely	11	0.02%	9	0.02%
Likely	30	0.05%	47	0.08%
Mainly	2	0.00%	5	0.01%
May	56	0.09%	76	0.13%
Maybe	3	0.00%	1	0.00%
Might	29	0.04%	10	0.02%
Mostly	2	0.00%	0	0.00%
Often	11	0.02%	26	0.04%

Ought	1	0.00%	0	0.00%
Perhaps	21	0.03%	16	0.03%
Possible + Possibly	26	0.04%	14	0.02%
Presumable + Presumably	2	0.00%	1	0.00%
Quite	8	0.01%	25	0.04%
Rather	22	0.03%	21	0.04%
Relatively	8	0.01%	2	0.00%
Roughly	1	0.00%	1	0.00%
Seem + Seems + Seemed + Seemingly	36	0.06%	41	0.07%
Should	82	0.13%	98	0.17%
Sometimes	1	0.00%	9	0.02%
Somewhat	3	0.00%	6	0.01%
Suggest + Suggests + Suggested	7	0.01%	9	0.02%
Suppose + Supposes + Supposed	7	0.01%	0	0.00%
Suspect	4	0.01%	2	0.00%
Tend to + Tended to + Tends to	2	0.00%	4	0.01%
Think + Thinks + Thought	27	0.04%	26	0.04%
Typical + Typically	3	0.00%	1	0.00%
Uncertain + Uncertainly	1	0.00%	6	0.01%
Unclear + Unclearly	0	0.00%	6	0.01%
Unlike + Unlikely	10	0.02%	10	0.02%
Usually	17	0.03%	12	0.02%
Would	70	0.11%	173	0.30%
Total	863	1.32%	1031	1.76%
Use of Boosters				
Actual + Actually	24	0.04%	38	0.06%
Always	16	0.02%	28	0.05%
Certainly + Certain	7	0.01%	19	0.03%
Clearly + Clear	43	0.07%	12	0.02%
Definitely + Definite	0	0.00%	2	0.00%
Demonstrate + Demonstrates + Demonstrated	6	0.01%	0	0.00%
Establish + Established	5	0.01%	11	0.02%
Evidently + Evident	4	0.01%	3	0.01%
Find + Finds + Found	30	0.05%	41	0.07%
Highly + High	60	0.09%	15	0.03%
In fact	21	0.03%	9	0.02%
Indeed	12	0.02%	9	0.02%
Know + Known + Knows	32	0.05%	57	0.10%
Must	17	0.03%	56	0.10%
Never	25	0.04%	35	0.06%
No doubt	4	0.01%	4	0.01%
Of course	14	0.02%	17	0.03%
Prove + Proves + Proved	5	0.01%	8	0.01%
Realize + Realized + Realizes	5	0.01%	9	0.02%

Really	10	0.02%	14	0.02%
Show + showed + shown + shows	52	0.08%	36	0.06%
Sure + Surely	11	0.02%	4	0.01%
True + Truly	14	0.02%	18	0.03%
Undoubtedly	2	0.00%	15	0.03%
Total	419	0.64%	460	0.79%
Use of Attitude Makers				
Admittedly	1	0.00%	1	0.00%
Agree + Agreeable + Agrees + Agreed	8	0.01%	6	0.01%
Amazed + Amazing + Amazingly	0	0.00%	2	0.00%
Correctly	1	0.00%	1	0.00%
Curious + curiously	0	0.00%	2	0.00%
Disagree + Disagreed + Disagrees	0	0.00%	1	0.00%
Disappointed + disappointment + disappointing + disappointingly	0	0.00%	1	0.00%
Dramatic + Dramatically	1	0.00%	1	0.00%
Essential + Essentially	16	0.02%	11	0.02%
Expected + Expectedly	14	0.02%	19	0.03%
Fortunate + Fortunately	0	0.00%	1	0.00%
Hopefully + Hopeful	0	0.00%	5	0.01%
Important + Importantly	36	0.06%	24	0.04%
Inappropriate + inappropriately	1	0.00%	0	0.00%
Interesting + Interestingly	3	0.00%	5	0.01%
Prefer/ably/able + Prefers	1	0.00%	2	0.00%
Remarkable + Remarkably	3	0.00%	1	0.00%
Shock + Shocked + Shocking + Shockingly	10	0.02%	9	0.02%
Striking + Strikingly	0	0.00%	3	0.01%
Surprise + Surprised + Surprising + Surprisingly	9	0.01%	14	0.02%
Understandable + Understandably	1	0.00%	3	0.01%
Unexpected + Unexpectedly	1	0.00%	2	0.00%
Unfortunate + Unfortunately	8	0.01%	12	0.02%
Unusual + Unusually	8	0.01%	0	0.00%
Total	122	0.19%	126	0.22%

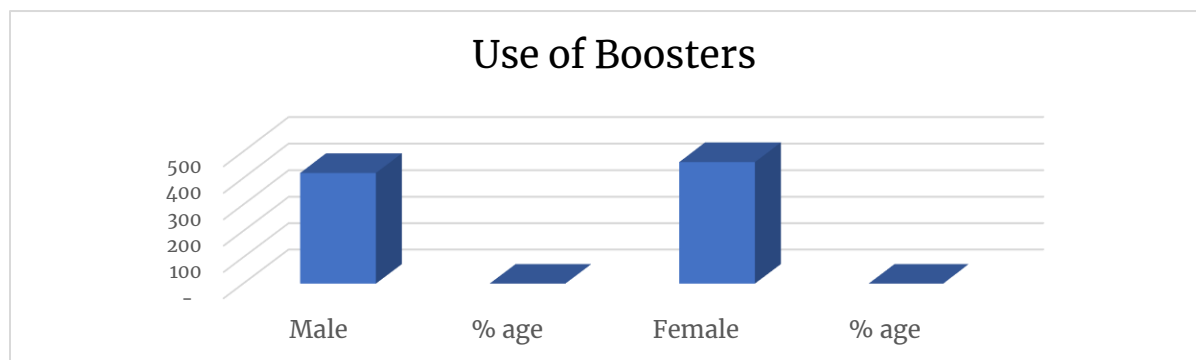


Figure 1: Graphical Representation of Use of Boosters

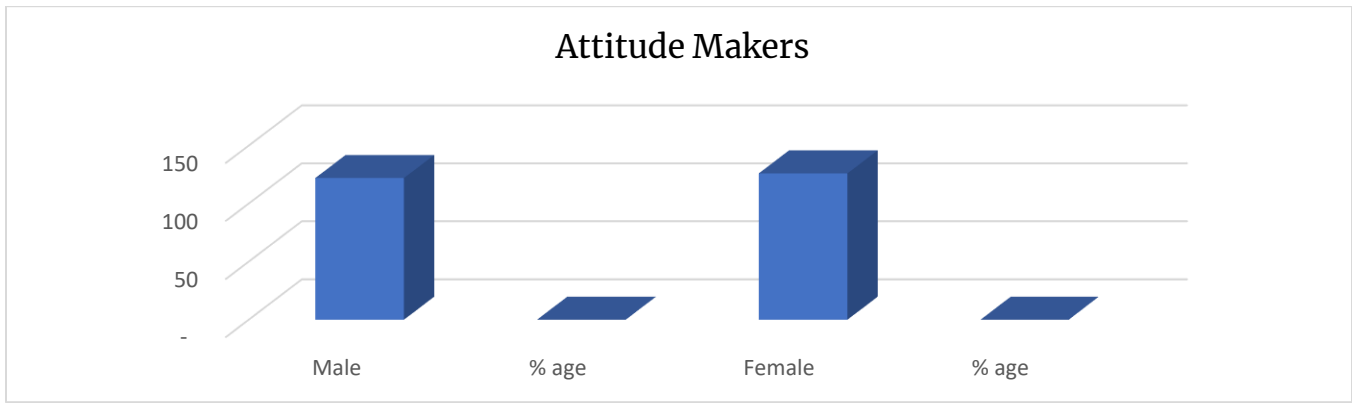


Figure 2: Graphical Representation of Use of Attitude Makers

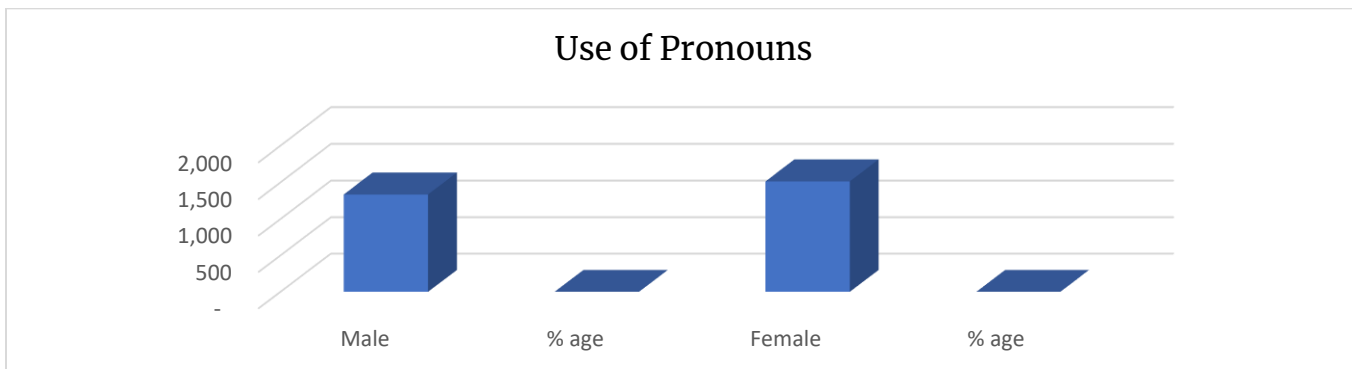


Figure 3: Graphical Representation of Use of Pronouns

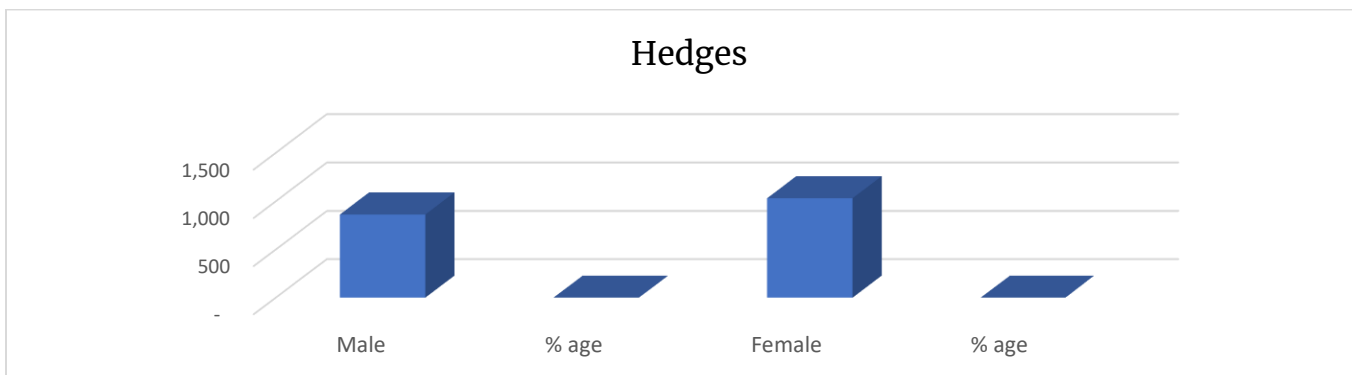


Figure 4: Graphical Representation of Use of Hedges

Discussion

How frequently female and male bloggers use MDMs in their writings in the leading newspaper of Pakistan is the main research question that has been caused and the researcher did a frequency count in the two gendered corpora. The instances MDMs of each category which includes boosters,

hedges, use of pronouns, and attitude makers; have been traced, coded, and counted to assess their distribution in each corpus and finally, these instances are compared between the two corpora (referred to Figures 1, 2, 3 & 4). The following table (Table 3) reveals the types, frequency, and percentage of the MDMs employed by male and female bloggers.

Table 3. Frequency and Proportion of MDMs Used by Male and Female Bloggers

Types of MDMs	M-Corpus	% Age	F-Corpus	% Age
Use of Pronouns	1,335	2.0%	1,513	2.6%
Hedges	863	1.3%	1,031	1.8%
Use of Boosters	419	0.6%	460	0.8%
Attitude Makers	122	0.2%	126	0.2%

As Table 3 illustrates, the frequency of all key types of MDMs (use of pronouns, attitude makers, boosters, and hedges) have been measured in the male and female corpora and the point to note that the use of these MDMs is more frequently employed by the female bloggers. This implied that the female gender is implied towards using these MDMs in their writings. In the female corpus, the frequency to use pronouns in their writings is 1513 which counts as 2.6% when it is compared with the male corpus, the use of pronouns is 1335 which counts as 2.0% in their writings. It is interesting to note that in the use of the first-person pronoun 'I' as in the case with male corpus, the frequency of 'I' is 99 that counts for 0.15% in the writings whereas, in the female corpus, the frequency is 173 that counts for 0.30% in the female corpus which is approximately double the number in comparison to the male use of the personal pronoun 'I'. Another striking feature has been noted with the use of second-person pronouns such as in the case of 'you' and 'your', then in the male corpus the frequency of these words is 125 concordances that count as 0.19% of the male corpus whereas in female corpus the frequency has been noted forty-seven times that counts as 0.08%. This infers those male bloggers are more inclined towards using second-person pronouns and less conscious about projecting themselves. It reflects that people with high status are more inclusive as was supported by previous research such as (Khan & Afsar, 2019) study of language in Pakistani blogs also noted that men used more inclined towards using collective personal pronouns.

Likewise, in the use of hedges in the male corpus the frequency of hedges is noted as 863

concordances that count for 1.3% whereas in the female corpus, it is relatively higher than the male corpus counts as 1,031 concordances, and in percentages, it is 1.8%. However, certain discrepancies have been noted when it comes to comparing certain individual instances of hedges such as 'argue' 'almost' 'around' 'broadly' 'doubt' 'estimate' 'fair' 'largely' frequently 'largely' 'maybe' 'might' 'mostly' 'possible' 'perhaps' 'presumable' 'rather' 'relative' 'suspect' 'suppose' 'think' 'typical' 'usual', then in the male corpus, these hedges are used more frequently than in female corpus. A similar case was with the use of boosters, in the male corpus, the frequency of these words is 419 concordances which counts as 0.6% of the whole male corpus whereas in female writings the use of boosters is relatively higher as 460 concordances count as 0.8%. Similarly, there are certain individual cases of booster words that have been counted more in the male corpus such as 'clear', 'high', 'in fact', 'indeed', 'show', and 'surely' and all these words refer to those male writers were assertive in presenting their arguments than female bloggers as supported by the previous research. Finally, in the case of attitude makers, though in the female corpus, it is counted slightly higher when it is converted in percentages then the results are the same i.e., 0.2%. It reflects that both genders are equally conscious when it comes to presenting their stance through writing. Attitude makers such as 'agree', 'essential', 'important', 'inappropriate', and 'shock', are more frequently used in the male corpus than in the female corpus. For a better understanding of these frequencies, a graphical representation has been given in Figure 5 appended on the next page.

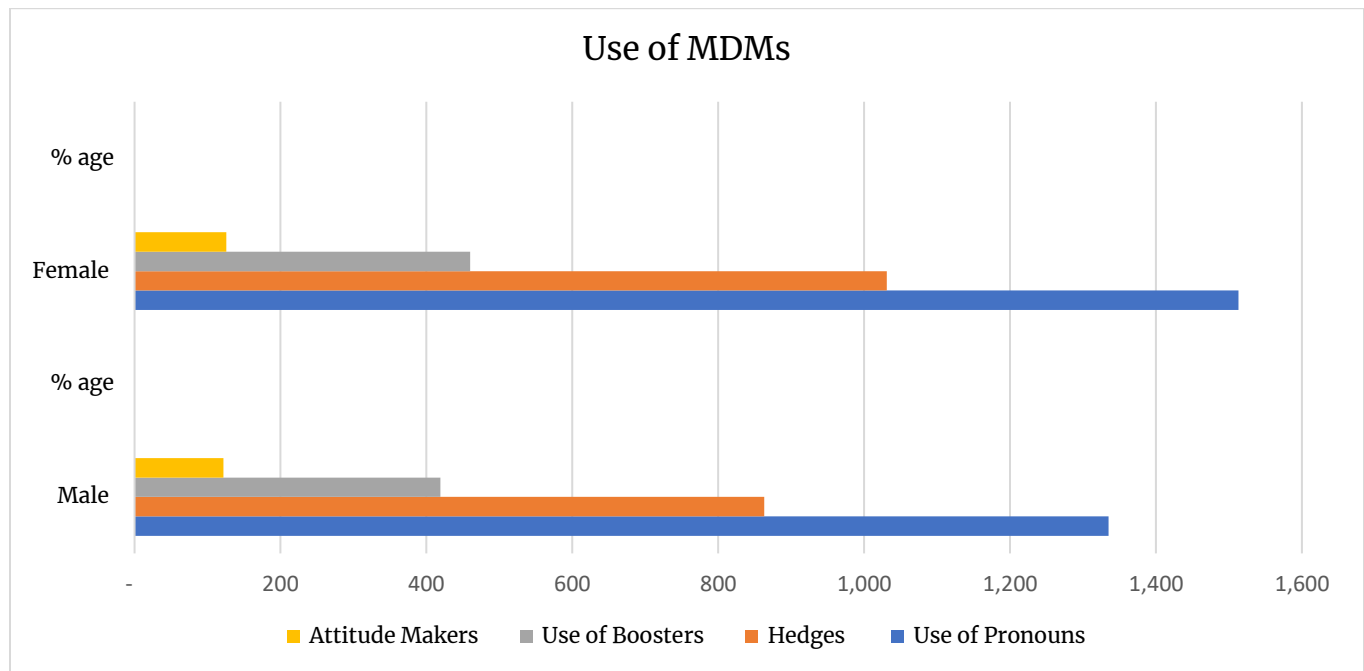


Figure 5: Graphical Representation of Use of Interactional MDMs

Conclusion

This study aims to explore the frequently used MDMs in the writings of male and female blogs by comparing the two corpora through the framework of Hyland (2005)'s framework using the methodology of corpus linguistics. For this analysis, two gender corpora have been built. The study finds that female bloggers are more inclined towards using these interactional metadiscourse markers in their writings that include hedges, pronouns, attitude makers, and boosters. This research focuses on one aspect and language can be studied and it needs to be studied from various perspectives within Pakistan to better understand linguistics nuances through different prisms as pointed out by Coates and Johnson (2001) that the study of language allows a uniquely "social" viewpoint on the study of gender differences (The New Handbook of Language and Social Psychology | Wiley, n.d.). The language of male and female bloggers continues to be a rich source of data offering critical insights and providing essential readings to the researchers working in the field of language and gender.

References

- Ädel, A. (2006). *Metadiscourse in L1 and L2 English*. Van Haren Publishing.
- Androutsopoulos, J., & Beißwenger, M. (2008). Introduction: Data and Methods in Computer-Mediated Discourse Analysis. *Language@Internet*, 5(2). <http://nbn-resolving.de/urn:nbn:de:0009-7-16090>
- Anthony, L. (2017). AntPConc (Version 1.2.1) [Computer Software]. (1.2.1) [Computer Software]. Waseda University. <https://www.laurenceanthony.net/software/antpconc/>
- Baker, P (2014). *Using Corpora to Analyze Gender*. Bloomsbury Publishing. <https://www.bloomsbury.com/uk/using-corpora-to-analyze-gender-9781441108777/>
- Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8, 243-257.
- Biber, D. E., & Burges, J. (2000). Historical Change in the Language Use of Women and Men: Gender Differences in Dramatic Dialogue. *Journal of English Linguistics*, 28(1), 21-37. <https://doi.org/10.1177/00754240022004857>

- Coates, J. (2004). *Women, men and language* (3rd ed.). Harlow: Longman.
- Hyland, K. (2004). Disciplinary interactions: metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing*, 13(2), 133–151. <https://doi.org/10.1016/j.jslw.2004.02.001>
- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173–192. <https://doi.org/10.1177/1461445605050365>
- K. Hyland (2000). *Disciplinary discourses: Social interactions in academic writing*. Longman, London.
- Khan, I. H., & Afsar, A. (2020). *A Corpus-Based Analysis of Differences in Language Use by Men and Women Bloggers of English E-Newspapers in Pakistan*. [Doctoral dissertation, International Islamic University, Islamabad]. Prr HEC. <http://pr.hec.gov.pk/jspui/handle/123456789/14436>
- Lakoff, R. (1973). Language and woman's place. *Language in Society*, 2(1), 45–79. <https://doi.org/10.1017/S0047404500000051>
- M.A.K. Halliday. (1994). *An introduction to functional grammar* (2nd ed.), Edward Arnold, London.
- 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. <https://doi.org/10.1080/01638530802073712>
- Schmid, H.-J. (2000). Do women and men really live in different cultures? Evidence from the BNC.
- Speer, S. A. (2005). *Gender talk: Feminism, discourse and conversation analysis*. London: Routledge.
- Talbot, M. M. (2010). *Language and gender* (2nd ed). Polity Press.
- The New Handbook of Language and Social Psychology | Wiley. (n.d.). Wiley.Com. <https://www.wiley.com/en-pk/The+New+Handbook+of+Language+and+Social+Psychology-p-9780471490968>