

A CORPUS-BASED SOCIOLINGUISTIC STUDY OF GENDERED LANGUAGE IN PAKISTANI JOURNALISM: THE FEMALE PERSPECTIVE

Tooba Maryam

*MPhil English (Linguistics), Department of English Language and Linguistics,
University of Sargodha*

Email: toobamaryam92@gmail.com

Prof. Dr. Ijaz Asghar

*Chairperson, Department of English Language and Linguistics, University of
Sargodha*

Email: ijaz.asghar@uos.edu.pk

Prof. Dr. Hafiz Ahmad Bilal

Higher Education Department, Punjab

Email: ahmadbilal.uos@gmail.com

Dr. Naveed Nawaz Ahmad (Corresponding Author)

*Assistant Professor, Department of English Language and Linguistics, University of
Sargodha*

Email: naveed.nawaz@uos.edu.pk

Abstract

This study investigates gender-based lexical choices in Pakistani News Discourse (2023–2025) through Corpus-Based Sociolinguistic Studies. Newspapers in Pakistan provide linguistic spaces where gender-based identities are generated, expressed, reinforced, and challenged. A purpose-built corpus of editorials and feature articles written by female journalists is used as the foundation to uncover the reflection and reproduction of gendered language patterns which demonstrate sociocultural power relations. Both quantitative and qualitative analysis of corpus is done for keyword and collocational patterns, recurrent lexical preferences and expressions that shape gendered representation. The patterns of gendered language in the writings of female journalists have been explored through methodological framework of Corpus Assisted Sociolinguistic Studies. The corpus analysis revealed that female journalists in Pakistani English newspapers prominently use gender-related lexical items, modal verbs, and collocational patterns to construct and assert their gendered identity. Key thematic domains included gender, education, politics, media, and culture, highlighting both advocacy and empowerment. Their language reflects agency, solidarity, and social awareness, while challenging traditional gender norms and stereotypes.

Keywords: *Corpus Assisted Sociolinguistic Studies, gendered language, editorials, collocations*

1. Introduction

Pakistani society is dynamic and complex where people with various cultural and linguistic backgrounds live together (Seifi, 2015). Like any other society, language in Pakistan is not only a means of communication but also a medium of expression of identity (Khalid, Nazir, & Jehanzeb, 2025). As a result of colonialism, the status of English in Pakistan is quite complex. On one hand, it is a language of education, modernity, and global connectivity and on the other hand it is a language of colonial legacy which leads to linguistic discrimination on the basis of class (Rahman, 2009). In this context, English language media plays a crucial role in expression and negotiation of gender based identity (Rehman & Zafar, 2025). The aim of this study is to explore how female journalists in Pakistani English newspaper discourse employ lexical choices and patterns of linguistic variation to negotiate and express their identity and to align with or challenge the established social norms.

Pakistani society is dominated by patriarchy, where women have to face various challenges such as gender discrimination and rigid societal norms (Rahman, 2002). Regardless, the constructive role played by Pakistani women in the establishment of progressive society and development of society

cannot be denied. Women stood side by side with men during the independence movement and since the creation of Pakistan have been contributing significantly to the development of Pakistan despite facing many challenges. Pakistani women express their opinions, views, suggestions and criticism through their writings in media discourse. How they create their gendered voices, negotiate their identities, which themes they deal with, what lexical choices they make in order to express those themes have been explored in this research.

1.1 Context and Background

Pakistani English is influenced by Urdu, regional varieties, and socio-religious norms. English language in Pakistan is considered the language of the elite and educated individuals; it is linked with administration, education, government, law, and legislation (Rahman, 1996). Due to its colonial legacy, English enjoys a prestigious position in society. Still, in a patriarchal society like Pakistan, the representation of women journalists and the language used by them remain underexplored.

Men and women tend to employ different discursive strategies and lexical styles shaped by societal expectations. In a society like Pakistan, gender roles are intertwined with religion, class, and tradition the examination of gendered language provides a deep socio-political insight (BMC Women's Health, 2022). For instance, headlines such as "Brave girl stands up to harassment" or "Female journalist breaks barriers" may appear progressive but simultaneously enforce gendered binaries which portray women as exceptions. This implies that since women are brave only in exceptional cases, an average Pakistani woman is not as brave as an average Pakistani man. Moreover, focusing on this aspect of discourse an illustration can be made showing how Pakistani media discourse negotiates power dynamics based on gender. Hence, a corpus-assisted analysis enables a systematic examination of lexical choices across a large dataset.

1.2. Gender, Media, and Language in Pakistan

The last decade has witnessed increased participation of females in political commentary, journalism and socio-cultural debates. Yet, we still can observe a strong reflection of patriarchal influence in our media discourse. The Pakistani female journalists are more vulnerable because they not only face safety risks but also they face gender discrimination in their recruitment and pay-scale (Jamil, 2020). In last few decades, Pakistan has observed multiple waves of feminism. Feministic movements such as "Aurat March" and campaigns such as "Me Too Pakistan" have given rise to both formal and informal debates about political, social and religious roles of women in our society. The public discussions on gendered based ideologies have become more critical many of which find expression through English language media outlets that target educated audiences. Consequently, analysing news discourse during 2023–2025 captures a transformative moment in Pakistani gender politics, as traditional roles and contemporary activism collide.

In this context, studying lexical choices provides insight into how journalists linguistically construct gendered identities either through empowerment oriented vocabulary or through traditional lexical framing. This study, therefore, indicate both ideological stance and changing gender discourse in Pakistan.

1.3. Research Gap and Significance

While there is extensive research conducted globally on gender and media representation, corpus-based studies focusing on Pakistani English are still limited. Moreover, most of the research conducted on gendered media discourse study how women are portrayed in media, but very few of them focus on how women themselves construct their identity through conscious lexical choices and thematic engagement. The main aim of this study is to explore, identify and recognise patterns of gendered language choices through a detailed corpus-based analytical methodology.

This research addresses that gap by constructing a specialised Pakistani English News Corpus (PENC 2023–2025), encompassing editorials and features authored by female journalists. Its purpose is to statistically analyse and recognise the patterns of lexical choices for identity negotiation and construction. The study thereby extends the discussion of World Englishes beyond linguistic description to include sociolinguistic and gendered identity formation.

2. Literature Review

The relationship between language and gender is complex and dynamic (Cameron, 2005). Gender, distinct from biological sex, is constructed in a society through established norms, beliefs, rules and ideologies and finds its expression through many factors and language is one of them (Eckert & McConnell-Ginet, 2003). Language reflects and constructs gender identities based on social norms and power relations. Media discourse deals with analysis of language used in news, advertisement, social media etc. Recently, Corpus based studies in sociolinguistics has made the research in this field more empirical and objective. It provides evidence to prove theories and reveal detailed pattern of linguistic variation.

For years, Sociolinguistics has dealt with the study of language and gender to understand how language is used differently by men and women and how these differences are a reflection of their social roles. Researchers have also tried to uncover whether these linguistic differences are related to biological differences. One of the first seminal works in this area included Lakoff's *Language and woman's place* (1973), in which she suggested that women's speech patterns are due to their social conditioning and not due to biological factors. She proposed that women's speech is more polite and less assertive because society expects them to sound soft and indirect. So, according to her, women's speech was shaped by social inequality. But her book was criticized by scholars because it implied women's speech to be inferior.

Later studies by Coates (2004), Holmes (1995), and Cameron (2005) further explained and elaborated the idea that gendered speech is not biological but social; it depends on how people express their identities through language. Spender (1980) and Fishman (1978) proposed dominance modal, according to which patterns of linguistic variation in women's speech are due to male dominance and patriarchy in society. In response to these early arguments, Deborah Tannen (1993) introduced the difference modal. According to her, men and women belong to different subcultures, which is the main reason for difference in their speech.

The paradigm later shifted from these initial works to social constructionism, which is the foundation of most contemporary sociolinguistic research related to gender and language. Gender was redefined as a performed identity; something that we negotiate and perform not something that we actually are (Butler, 1990). This idea was further elaborated by Eckert & McConnell-Ginet (1992), through their concept of language as community practice. People use language to stabilize their positions as members of a society. Bucholtz and Hall (2005) proposed that identity is constructed step by step through discourse. Over the last few decades, Pakistani researchers have increasingly examined the links between language, ideology, and gender in media discourse. One significant contribution in this area is by Asghar and Akhter (2022), who conducted a corpus-based Critical Discourse Analysis of three major English newspapers in Pakistan. Their findings show that women's voices are both linguistically reduced and structurally marginalized. News discourse in Pakistan is dominated by patriarchal ideologies.

Gender ideologies appear consistently in media across various genres. Haris (2024), for instance, employed multimodal discourse analysis to examine English-language magazine advertisements. The results of the research indicated that advertisements meant for women were mostly dominated by lexical items related to beauty, home-making, and care-giving while those targeting men emphasize competence, authority, and decision-making. Such linguistic choices reinforce traditional gender norms.

Gender representation has also been studied in the educational context as well. Mahmood, Kaur, and Daghigh (2021) analyzed English-language textbooks used in Pakistan through a Critical Discourse Analysis framework. Their study shows that female characters are often portrayed in passive, supportive, or domestic roles, whereas male characters are positioned as leaders and decision-makers. A similar pattern was observed by Gul and Durr-e-Nayab (2025), who reviewed English-medium school textbooks in Punjab and found structured ideological biases embedded in both lexical and syntactic features.

Women in Pakistan feel more pressured because they are at higher risk of being harassed. Yasmeeen (2025) conducted a research to investigate online harassment of female politicians and uncovered that gendered insults and moral attacks act as tools of suppression that are used to weaken the voice of women in public spheres. This immoral social pressure plays a devastating role that marginalizes women in Pakistan.

These findings align with long-standing sociolinguistic observations regarding gendered discourse practices and further demonstrate how linguistic patterns reflect and reproduce gendered expectations in Pakistan. Collectively, these studies show that gendered language in Pakistani media, across various genres including news, textbooks, advertisements, and online platforms, is shaped by societal norms and ideological structures. This research directly supports this ongoing present study, which adopts a corpus-based sociolinguistics approach to examine how female journalists construct, negotiate, and establish their gendered identities through lexical choices and thematic engagement.

3. Research Objectives

- To identify gender-based lexical patterns in the writings of female journalists' through collocation and concordance analysis.
- To interpret the linguistic choices of female journalists through a sociolinguistic lens, with particular attention to stance-taking and identity construction.
- To examine how journalists' lexical choices reflect, reinforce, or challenge sociocultural gender ideologies in Pakistani media discourse.

4. Research Questions

- What gender-based lexical choices are predominant the in the writing of female journalists in Pakistani English newspapers (2023–2025)?
- How do the collocational patterns in female journalists' writings reflect their gendered identity?
- Which themes are most prevalent in female journalists' writings and what does their stance reveal about their gendered voices?
- How do female journalists in Pakistan construct their gendered identity through lexical choices and selected themes?

5. Methodology

5.1 Research Design

This study applies a corpus-based sociolinguistic design to investigate the lexical choices, collocations and keywords used by female journalists so that patterns of gender-based variations in Pakistani English journalism can be uncovered. The aim of this study is to demonstrate how these patterns of gender-based lexical choices play a key role in the construction and expression of gender-based identity. A corpus consisting 30,000 words has been constructed. The corpus contains almost 35 articles written by female journalists in Pakistani newspapers from 2023-2025. The articles are selected through simple random sampling.

5.2. Data Collection and Analysis

The data is collected from three major English Language Newspapers in Pakistan Dawn, The Express Tribune and The News International. The corpus covers the period from January 2023 to November 2025. The gender of the author is identified author through biographies, Official journalist profiles and Professional networking platforms. Ambiguous authors have been avoided.

5.3. Analytical Framework

Corpus tool Antconc has been used to analyze data efficiently. Concordance analysis has been done to reveal key terms used most frequently by female journalists. These key terms are in turn used to uncover major themes of these articles. Collocates of key terms have been studied to identify gender-specific patterns to reveal various themes. The quantitative patterns revealed through corpus analysis are interpreted through sociolinguistics concepts, in order to uncover how these patterns reflect professional identity and social positioning.

6. Findings

The concordance analysis of corpus has revealed frequent key terms used by female writers. The most frequent content words are related to domains of gender, society, politics, education, media, language, society and culture.

6.1 Gender-related Terms

Rank	Token	Frequency
1	Women	177
2	Men	35
3	Woman	25
4	Female	17
5	Transgender	16

The randomly selected corpus is dominated by the word “women” (177), which shows that Pakistani female writers write about women as women. Traditionally, due to the dominance of patriarchy, women have been underrepresented in news discourse. Pakistani women are now challenging and countering this, as evident in the corpus. The word “Men” (35) occurs approximately four times less frequently indicating that male-centred narratives are secondary. This indicates the active participation of women in shaping gender discourse rather than becoming passive subjects of it. This reflects the indexicality of gendered identity; language being used to signal belonging and stance (Eckert, 2012). In short, this lexical cluster reveals that female journalists use language as a form of gendered self-representation

The appearance of the word “Transgender” (16) indicates the awareness of gender-diversity and inclusivity. Transgender Community in Pakistan has been facing extreme marginalization and even speaking about them has been considered taboo. Collocates of the word “Transgender” in the corpus are (act, movements, rights, protection, strengthened) indicating the problems faced by transgender community and a need to empower them. These lexical choices show that Pakistani women demonstrate commitment to inclusivity and to breaking traditional binary gender construction.

6.2 Society and Politics

Rank	Token	Frequency
1	Government	44
2	Power	41
3	Political	38
4	Policy	26
5	Rights	25

This lexical cluster reveals that Pakistani female journalists actively engage in themes of administration, policy-making, and power relations. Their writing occupies public and political

space, which counters the stereotypes that women's writing is personal and emotional. Women express their opinions and insights about major issues of the state, challenging the traditional gender norms, establishing new standards, and redefining their role in society.

The frequent mention of words like "rights, policy, government" connects this discourse to reform and advocacy, indicating politically conscious use of language. It shows civic responsibility and solidarity, which aligns with Holmes' (2013) sociolinguistic findings that women use language as a tool for inclusion and equity. These terms also indicate that women are well aware of unequal power relations and are actively expressing them through language. In short, these terms position female journalists as politically literate and socially responsive agents, actively engaging with governance and justice through their lexical choices.

6.3 Education, Media and Language

Rank	Token	Frequency
1	Education	56
2	Media	60
3	Language	49
4	English	31
5	Journalists	26
6	Urdu	19

The terms related to language, media, and education occur on average more frequently than the terms relating to governance, and advocacy indicating a more prominent role and interest of women in education and language. Collocates of education in the corpus are social change, foundation, children, development, opportunities, and English, giving a positive connotation to the word. The role of women in the sector of education in Pakistan is undeniably central and this is evident through discourse created by women. Female writers reflect on the role of language and education in shaping society. The occurrence of the word English with education points towards the fact that English is the medium of instruction in most educational institutions in Pakistan. The word English occurs 31 times, while the word Urdu appears 19 times in the corpus negotiating between global prestige and local identity (Rahman, 2012; Mahboob, 2009). This data suggests that Pakistani female journalists are meta-linguistically aware and have socio-educational consciousness. The inclusion of term journalists (26) also show that female writers actively talk about their profession and its challenges. Thus, this cluster constructs linguistically self-aware, intelligent and educated female identity.

6.4 Socio-cultural Domain

Rank	Token	Frequency
1	Children	37
2	Social	32
3	Public	31
4	Society	27

Rank	Token	Frequency
5	Culture	12

This data suggest the human-centered and community-oriented aspect of women’s writing. Frequent references to children, social, and society highlight empathy, solidarity, and a sense of community making these features a crucial part of females’ identity. The presence of terms like public and culture show broader social and cultural framing of identity emphasizing shared experiences and collective narrative. Female journalists employ an empathetic narrative style, combining personal concerns with collective responsibility.

6.5 Modality and Voice

Rank	Token	Frequency
1	Will	80
2	Can	60
3	Would	53
4	May	48
5	Must	16

Modal verbs reveal a lot about assertion, authority, confidence, and stance of the author. The most frequent modal verb in the corpus is ‘will’ (80) which indicates a strong future orientation in discourse authored by female writers. Their writing style is directive yet optimistic, showing belief in the possibility of reform. Employment of the word ‘will’ shows a strong predictive voice and a deep thorough understanding of events and their future consequences. “Can” is the second most frequent modal verb in the corpus. ‘Can’ is used to show possibility and ability. In the corpus, it is used both positively as well as negatively. In the negative sense it is used with the words like ‘happen’ and ‘misogyny’ suggesting that what is happening with victimized women can happen with anyone among us and misogyny can’t be removed from the minds of those individuals in which it has already been deeply ingrained. Such as ‘it can happen to us’ and ‘we can imagine what they will do’. In the positive sense, this word is used to show ability and possibility. Such as ‘Policies must include contributions women can make in stabilising Pakistan’ and ‘Women in management positions can have a huge impact on their female colleagues’. It shows that women in written discourse make linguistic choices to convey a sense of group belonging, to point toward the challenges faced by women in Pakistan and to emphasize on empowerment of women by pointing out their abilities. They construct a female identity of resilience and empowerment.

7. Conclusion

The aim of this study was to explore how female journalists in Pakistan make lexical choices to construct their gendered identity and how they position themselves in society. A corpus was built based on news articles and features written by female authors published during the period from 2023 to 2025. A detailed corpus analysis was conducted through concordances and data obtained was explained through thematic analysis. The most frequently occurring words, excluding function words, included Pakistan (214), women (177), power (41), education (56), media (60), language (49), and rights (25). These words suggest a strong thematic engagement with domains like gender, society, politics, education, media, language, and culture. The terms related to gender and a detailed key word in context (KWIC) study confirm that female journalists actively participate in constructing

gender identity and play an active role in representing women in news discourse where previously they were underrepresented. They highlight the challenges faced by women in Pakistani society as well as share the problems and discrimination faced by them in their profession. Moreover, high frequency of terms related to government, administration, politics, and civic rights prove their political awareness breaking the stereotypes regarding the fairer sex. The words related to education, culture and language were higher in frequency than the words relating to domain of politics which shows a more active participation of women in sector of education and language learning. The terms like 'Us', 'We', 'Children', 'Society' create a sense of community and solidarity. The frequency of modal verbs shows that their writing combines assertion (will, can) with responsibility (should, must) and consideration (would, may), forming a balanced discourse that reflects both empowerment and empathy.

References

- Arshad, N., & Yasmin, M. (2025). Gendered linguistic patterns in the online harassment of female politicians in Pakistan. *Journal of Applied Linguistics and TESOL*, 8(2). <https://doi.org/10.63878/jalt891>
- Asghar, S., & Akhter, U. (2022). Representation of women: A corpus-based analysis of Pakistani English newspapers. *Journal of Communication and Cultural Trends*, 4(2), 61–80. <https://doi.org/10.32350/jcct.42.04>
- BMC Women's Health. (2022). *Perpetuation of gender discrimination in Pakistani society: Results from a scoping review and qualitative study conducted in three provinces of Pakistan*, 22. <https://doi.org/10.1186/s12905-022-02011-6>
- Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse Studies*, 7(4–5), 585–614.
- Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. Routledge.
- Cameron, D. (2005). *Language, gender, and sexuality: Current issues and new directions*. *Applied Linguistics*, 26(4), 482–502.
- Coates, J. (2004). *Women, men and language: A sociolinguistic account of gender differences in language*. Routledge.
- Eckert, P., & McConnell-Ginet, S. (1992). Think practically and look locally: Language and gender as community-based practice. *Annual Review of Anthropology*, 21, 461–490.
- Eckert, P., & McConnell-Ginet, S. (2003). *Language and gender*. Cambridge University Press.
- Fishman, P. M. (1978). Interaction: The work women do. *Social Problems*, 25(4), 397–406.
- Haris, M. (2024). Linguistic dimensions of gender representation: A multimodal discourse analysis of Pakistani English magazine advertisements. *Policy Research Journal*, 2(4), 2292–2326.
- Jamil, S. (2020). Suffering in silence: The resilience of Pakistan's female journalists to combat sexual harassment, threats and discrimination. *Journalism Practice*, 14(2), 150–170. <https://doi.org/10.1080/17512786.2020.1725599>
- Khalid, A., Nazir, K., & Jehanzeb, G. (2025). Language, identity, and power: Sociolinguistic patterns in urban Pakistan. *ASSA Journal*, 4(02), 1660–1669. <https://www.assajournal.com/index.php/36/article/view/1095>
- Lakoff, R. (1973). Language and woman's place. *Language in Society*, 2(1), 45–79.
- Mahmood, S., Kaur, S., & Daghigh, A. J. (2021). Role of gender in Pakistani ELT textbooks: A critical discourse analysis approach. *Cultura & Lingua*, 2(1), 9–14. <https://doi.org/10.37301/culingua.v2i1.17>
- O'Neil, J. M., Good, G. E., & Holmes, S. (1995). Fifteen years of theory and research on men's gender role conflict: New paradigms for empirical research. In *Parts of this chapter were presented at the 102nd Annual Convention of the American Psychological Association, Los Angeles, CA, Aug 1994*. Basic Books/Hachette Book Group.
- Rahman, T. (1996). *Language and politics in Pakistan*. Oxford University Press.
- Rahman, T. (2009). Language ideology, identity and the commodification of language in the call centers of Pakistan. *Language in Society*, 38(2), 233–258.
- Seifi, P. (2015). Language policy in multilingual and multicultural Pakistan. *Advances in Social Sciences Research Journal*, 2(3), 156–163.
- Spender, D. (1980). *Man made language*. Routledge & Kegan Paul.
- Tannen, D. (Ed.). (1993). *Gender and conversational interaction*. Oxford University Press.