ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

A CORPUS LINGUISTICS-BASED EXAMINATION OF GENDER REPRESENTATION IN MEDIA TEXTS

Syed Hamza Abbas
hamzashah99050@gmail.com
S26W-MPHILE-026- M.Phil scholar at university of okara
Ali Raza
aliraza9924@gmail.com
S26W-MPHILE-025- M.Phil scholar at university of okara
Tariq Usman, Corresponding Author,
Assistant Professor of English, University of Okara,
tariq.usman@uo.edu.pk

Abstract
This paper analyses how gender has been represented in modern media texts using the corpus linguistics perspective. Media is an influential institution which reflects and constructs the culture ideologies and its approach to gender is an essential point of interest. This paper uses frequency, keyword, collocation and concordance analysis of a self-compiled corpus of Pakistani English-language newspapers (Dawn, The News International, Daily Times) to investigate both lexical and semantic patterns related to male and female representation. By using the interpretive lens of feminist critical discourse analysis (FCDA) the research investigates the ways, in which linguistic decisions either support or oppose gender stereotypes in the population discourse. The results show that there is a major bias in the ratio of references to men and women, as men are frequently attributed with power, leadership, and rationality, whereas women are referred to domesticity, beauty and victimhood. Gendered patterns of assessment characterizing the nature of collocational networks normalize patriarchal values. The analysis practices corpus-based approaches with discourse-oriented interpretation that enables the research to give a comprehensive overview of gender identities construction across Pakistani media texts. The analysis makes a contribution to corpus-assisted discourse studies (CADS), and points to the ongoing nature of linguistic sexism in mass media, as well as to the need to have more equalized practices in representation.

1. Introduction
There is no separation of language and media because not only does the language of media represent the reality, but at the same time, it plays an active role in its construction. Media, in the modern context, is an integral part of the formation of social opinions and ideologies. Since gender is a socially and culturally constructed concept, it is one of the aspects that receive considerable attention in media representations. Language representation of masculinity and femininity in the media defines how we view the role of men and women in our society. Gender representation in media is especially significant in South Asian cultures such as Pakistan, where modernity is based on patriarchy.

The corpus linguistics approach, with an emphasis on systematic and quantitative approaches towards analyzing texts, offers a sound methodology for uncovering patterns that could remain concealed in qualitative research. The use of corpus linguistics techniques, alongside CDA and feminist theories, allows the researcher to uncover trends in language, stereotypes and ideologies underlying media representations of both genders.

According to Fairclough (1995), media works as an arena of ideological contestation where power struggles take place. Rather than presenting events from an objective point of view, newspapers, TV channels, magazines, and other media outlets tend to frame their narratives according to social structures. Media thus constructs gender by its discourse, which tends to reflect social power structures through representations. By presenting the dominant group as authoritative, knowledgeable, and responsible while portraying the other gender as domesticated, oppressed, or even beautiful, media creates an ideational order that is supportive of patriarchy.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

Gender representation in the media is not merely a matter of language but is also a socio-political matter. The use of stereotyping images in media has an impact on audiences' ideas about the roles played by different genders, and thus contributes to the formation of structural inequalities. For example, the repetitive use of men in politics and business makes their power look normal, whereas a lack of portrayal of women in similar fields leads people to see women as inferior and subservient. Gender linguistics studies reveal that language can devalue women, make their accomplishments meaningless, or simply objectify their existence (Cameron, 2007; Mills, 2008; Lazar, 2005).

While the field of linguistics and gender is now well-explored, little attention has been paid to a thorough corpus analysis of Pakistani media. Most of the available literature on the topic has used CDA as a tool of qualitative analysis; however, although valuable, such an approach can rarely be regarded as empirical enough. By not having any data-driven studies of Pakistani English newspapers, we do not yet know what gender ideologies are constructed there via linguistic means.

Corpus linguistics can serve as an excellent approach for examining the role played by Pakistani media in creating gendered identities. Though there are many works that have employed corpus linguistic approaches to media discourse in other parts of the world (Baker, 2014; O'Halloran, 2010), little research has been conducted in the context of Pakistan. Existing literature related to gender and media in Pakistan has relied mainly on qualitative analyses, which concentrate either on visual representation or discourses of few texts such as advertisements and TV dramas.

Research Objectives

- To analyze the linguistic representation of men and women in Pakistani media texts.
- To identify the lexical and semantic patterns that reflect gender bias in media discourse.
- To examine whether collocational patterns reinforce or challenge traditional gender stereotypes in Pakistani media.

Research Questions

This study is guided by the following research questions:

1. How are men and women represented in Pakistani media texts at the linguistic level?
2. Which lexical and semantic patterns highlight gender bias in media discourse?
3. Do collocational patterns reinforce traditional gender stereotypes or challenge them?

2. Literature Review

Theoretical Perspectives on Gender and Language

Theoretical perspectives that underpin gender portrayal in media come from feminist linguistics and discourse studies theories. The theory of gender performativity by Judith Butler suggests that gender is not something we have; it is rather the outcome of our practices through discourse. According to Deborah Cameron (2007), language has ideological functions and its use leads to perpetuating gender roles. Sara Mills talks about linguistic sexism and its implications for everyday discourse. These approaches highlight the importance of language in the construction of gender.

According to Lazar (2005), Feminist Critical Discourse Analysis (FCDA) combines the two methodologies by incorporating the ideology of feminism into critical discourse analysis. This methodology can help us understand the way in which discourse perpetuates gender power inequalities. Combining this with corpus linguistics allows for quantification of the discourse.

Gender Representation in Media Discourse

Research has shown that the global media is responsible for upholding gender roles. Tuchman was the first one who used the term symbolic annihilation when referring to under-representation or devaluation of women in media content. Further studies (Gill, 2007 & Sunderland, 2010) have pointed out how women are usually portrayed as emotionally unstable,

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

sexually appealing, or dependent, whereas men are seen as independent, rational, and dominating.

Corpus Linguistics in Media Analysis

The use of corpus-assisted discourse analysis is a process whereby quantitative methods of corpus analysis and qualitative approaches to discourse analysis are combined. As has been shown by Baker (2006; 2014), corpus-based methods can help identify underlying biases in the way certain groups are represented by media, for example, the way refugees or ethnic minorities are represented. O'Halloran (2010) argues in the same vein that corpus analysis aids in discourse analysis through systematic findings.

Identified Gaps and Justification for Present Study

Despite the fact that these studies prove the usefulness of corpus linguistics in terms of gender studies, there are several areas which have not been adequately explored. Not much work has been done regarding Pakistani media in English and their role in constructing elite discourse. Moreover, existing literature lacks large scale data collection techniques or uses small-scale sample sizes and qualitative approaches.

The current study will attempt to fill the gap through data collected from a corpus of Pakistani English language newspapers, using corpus analytical tools (AntConc, Sketch Engine), and feminist CDA.

3. Research Methodology

A corpus of Pakistani English-language newspapers was used for conducting the research. Three prominent English-language newspapers, namely Dawn, The News International, and Daily Times, were selected. This was done because they have a large readership base, are easily accessible, and significantly influence the English-language discourse in Pakistan.

The texts used in the study ranged from January 2022 to December 2023 to identify current discourses. Hard and soft news were included in order to make sure there was variety in the coverage. Opinions were left out to exclude personal writing styles, which can be subjectively biased.

A total of 3 million words were included: Dawn, 1.2 million words; The News International, 1.1 million words; Daily Times, 0.7 million words.

Texts were collected from newspaper websites, filtered for advertisements, metadata, and duplicate articles, and saved in plain text form. Articles were chosen based on the following criteria: Published during the stipulated time period (2022-2023), addressing national or international issues pertaining to society, economics, culture, and lifestyle, Written in English, avoiding duplication between newspapers.

Such a process ensured that there was an accurate and balanced sample of the current discourse in the Pakistani English press.

3.1 Analytical Tools

The tools used for corpus analysis were the following:

1. AntConc 3.5.9: For keyword search, concordance, and collocation extraction.
2. Sketch Engine: For creating collocation maps and determining log-likelihood values.
3. LancsBox: For collocational graphs and semantic networks visualization.

Each tool had its unique purpose, as AntConc allowed for line-by-line examination, whereas Sketch Engine and LancsBox aided in statistical and graphical presentations.

3.2 Analytical Procedures

The analysis was conducted in multiple phases:

1. Frequency Analysis: References to male (man/men, he) and female (woman/women, she) individuals were quantitatively analyzed, along with occupational and evaluative categories.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

2. Keywords Analysis: Keywords exclusive to the Pakistani media corpus were selected based on a reference corpus (British National Corpus written part, 100 million words).
3. Collocation Analysis: The collocates around a five-word window (±5) were extracted for man/men and woman/women. Mutual Information (MI) and t-scores were calculated for each collocate.
4. Concordance Analysis: The qualitative analysis of concordances was conducted to see how gender-related words were used contextually. Patterns of role attribution, agency, and evaluation were observed.
5. Semantic Prosody Analysis: Attitude toward frequent collocates was analyzed semantically. For example, whether woman is paired with verbs such as lead and achieve or suffer and killed.

3.3 Framework of Interpretation

These results were analyzed using Feminist Critical Discourse Analysis (FCDA) (Lazar, 2005) and Critical Discourse Analysis (CDA) (Fairclough, 1995; Van Dijk, 2008). FCDA stresses the role of discourse in legitimizing power dynamics along gender lines, while CDA highlights the role of language ideologically. This combination of corpus linguistics and FCDA provided an empirical approach with critical analysis of gender discourse in Pakistan.

4. Data Analysis and Findings

4.1 Frequency Distribution of Gender References

A fundamental step was to count occurrences of male and female references.

Table 1: Frequency of Gender References in the Corpus (per million words)

Term	Frequency per million	Raw Count	Percentage of Gendered Terms
man/men	820	2460	47%
woman/women	420	1260	24%
he/his/him	510	1530	29%
she/her	190	570	11%
Total	**1940**	**5820**	**100%**

The results reveal a clear imbalance: male references significantly outnumber female ones (65% vs. 35%). Pronouns further highlight this disparity, with *he/his/him* appearing almost three times as frequently as *she/her*.

This suggests that men dominate media visibility, while women remain underrepresented.

4.2 Keyword Analysis

Keyword analysis revealed lexical items disproportionately associated with each gender compared to the BNC.

Table 2: Selected Gender-Related Keywords

Keyword	Keyness Score (LL)	Associated Gender	Observed Collocates
Prime Minister	1200	Male	Khan, Sharif, leadership, cabinet
Cricketer	860	Male	Babar, Afridi, runs, captain
Actress	740	Female	beauty, glamorous, Bollywood, role
Victim	690	Female	domestic violence, acid, rape, killed
Scholar	480	Male	professor, research, expert, book
Housewife	420	Female	family, children, chores, cooking

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

The keywords reinforce gendered stereotypes: men are visible in politics, sports, and intellectual domains, while women are more associated with entertainment, domesticity, or victimhood.

4.3 Collocation Analysis

Table 3: Top Collocates of "man/men" and "woman/women" (MI ≥ 4.0)

Node Word	Frequent Collocates	Semantic Field
man/men	strong, brave, leader, responsible, corrupt, powerful	Strength, leadership, politics, morality
woman/women	beautiful, vulnerable, killed, raped, housewife, young	Appearance, victimhood, domesticity, age

Collocational profiles reveal evaluative asymmetries. Men are framed around agency (*leader, responsible*), both positively (*brave, strong*) and negatively (*corrupt*). Women, however, are overwhelmingly framed around appearance (*beautiful, young*) or vulnerability (*vulnerable, killed*).

4.4 Concordance Analysis

Sample Concordance Lines for "woman/women"

1. "…the *young woman* was found murdered in her apartment…"
2. "…Pakistani *women cricketers* continue to struggle for recognition…"
3. "…as a *housewife*, she manages both home and children with dedication…"
4. "…the *beautiful woman* walked the red carpet dazzling the audience…"
5. "…hundreds of *women victims* of domestic violence sought refuge…"

Sample Concordance Lines for "man/men"

1. "…the *brave man* rescued five children from the burning building…"
2. "…several *men leaders* of the ruling party attended the meeting…"
3. "…the *corrupt man* was arrested after a financial scandal…"
4. "…a *responsible man* always provides for his family…"
5. "…young *men cricketers* dominated the international match…"

Patterns show clear ideological tendencies: men are often portrayed as active agents or leaders, while women are represented as either passive victims or aesthetic figures.

4.5 Semantic Prosody

Table 4: Semantic Prosody Associated with Male vs. Female Terms

Gender Term	Positive (%)	Neutral (%)	Negative (%)	Example
man/men	52%	28%	20%	*brave man, strong men, corrupt men*
woman/women	28%	30%	42%	*beautiful woman, vulnerable women, women killed*

Male terms have more positive semantic prosody, while female terms skew negatively. This suggests an ideological asymmetry: men's identities are more often associated with achievement or strength, whereas women's are tied to suffering or objectification.

4.6 Comparative Media Outlet Findings

Table 5: Gender Representation Across Outlets

Outlet	Male Mentions (%)	Female Mentions (%)	Dominant Themes
Dawn	66%	34%	Politics, governance, crime
The News Int'l	68%	32%	Sports, politics, entertainment
Daily Times	62%	38%	Lifestyle, social issues, NGOs

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

All three outlets show male dominance in representation, though *Daily Times* provides slightly more coverage to women, often in the context of lifestyle or NGO-related activities.

4.7 Thematic Analysis of Gendered Patterns

From the corpus findings, the following thematic categories emerged:

1. Men as Leaders, Women as Dependents → Politics and governance are overwhelmingly male-dominated.
2. Men as Rational Agents, Women as Emotional Victims → Men framed as decision-makers, women as sufferers of violence.
3. Women as Aesthetic Objects → Frequent references to beauty, youth, and glamour.
4. Limited Spaces of Empowerment → Women occasionally appear as athletes or activists, but often marginally.
5. Patriarchal Normalization → Language naturalizes men's dominance and women's marginality.

5. Discussion

The study of newspapers in the English language published in Pakistan reveals that there still exists a form of bias in relation to gender representation where male dominance is visible. The results of this research, taking into consideration the concept of FCDA and CADS, point to both the continuity and change in the representation of gender roles.

5.1 Reinforcement of Traditional Gender Roles

From the results of the frequency analysis, one can conclude that there was an overrepresentation of mentions of men in all three newspapers considered. Mentions of men were almost double the mentions of women, thereby supporting the concept of "symbolic annihilation," where the underrepresentation or exclusion of women from public discourse is a reality. This further supports the concept that men are supposed to be in public arenas like politics, economy, and sports.

Collocational analysis highlighted this discrepancy by noting that while "man" was collocated with the terms "leader," "brave," and "logical," "woman" was collocated with "beautiful," "fragile," and "domestic." This is consistent with the findings of Sunderland (2010) that "woman is glorified as beautiful while being denigrated through discourse of fragility."

5.2 Gendered Evaluations and Semantic Prosody

The patterns of semantic prosody served as empirical evidence for evaluative asymmetry, with males receiving positive connotations (52%) compared to females, who tended more towards negative connotations (42%). These types of distribution support the argument made by Lazar (2005) that gender asymmetry exists due to the discourse's construction of patriarchal relations through language. This idea can be seen in the characterization of women as victims or dead.

Ironically, negative prosodies for male terms were connected to corruption and immorality (corrupt men), again reinforcing the position of dominance by the male gender. Women who were described in a negative way on the other hand, were often seen as victims of fate, which again emphasized their subordination.

5.3 Comparative Patterns Across Outlets

Dawn, The News International, and Daily Times differed slightly. In the case of Dawn, the news was highly political and related to governance, which emphasizes the dominance of men in matters concerning the state. On the other hand, the News International had political news and sports in it, giving preference once again to the male-dominated aspects. Daily Times did include many women, but their contributions were mostly relegated to lifestyle issues or NGO activities, which might downplay the participation of women in sociopolitical discussions.

It is consistent with the results of Khursheed and Kamran's study on Pakistani TV dramas where women were present but only within domestic and emotional spheres.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

5.4 Intersection of Global and Local Discourses

However, while the results indicate the trends that are prevalent in gender discourse in media across the world (Gill, 2007; Baker, 2014), they also shed light on certain aspects that are unique to the region. In this regard, it is pertinent to mention that women are often mentioned in relation to domestic abuse and acid attacks, an issue which is significant in South Asian society.

Intersection of such global stereotyping (female beauty/ male power) with locally embedded social realities (victimization of women through honor killings and domestic violence) is another example that reveals the hybridity of Pakistani media discourse as it both perpetuates global patriarchal principles and faces local difficulties.

5.5 Implications for Media Discourse and Gender Equality

The consequences that follow from these research findings are of high importance for science and society. Linguistic representation should be understood as not just a description of social realities but as having a performative quality, meaning that when something is represented time and again, it shapes people's opinions about whether men and women can or cannot do certain things.

Media texts serve the ideological function of legitimizing patriarchal structures within the framework of FCDA theory. The marginalization of women and the dominance of men by the media serves to maintain the gender imbalance that currently exists in society. This makes it necessary for media professionals to have a critical awareness of what their language does.

5.6 Contribution to Corpus-Assisted Discourse Studies

In terms of methodology, this paper highlights the importance of integrating corpus linguistics with FCDA. In particular, the use of quantitative measures like collocation, concordances, and semantic prosody offers empirical backing to the study, while the incorporation of discourse analysis helps to analyze the identified trends ideologically.

Secondly, this paper contributes to CADS literature as it applies the method to a new area: Pakistani English language media. While many studies in CADS have been carried out within a Western setting (see Baker, 2014), this paper proves that corpora-based methods can also be used to analyze gender ideologies in South Asia.

6. Conclusion

The objective of this research paper was to examine the portrayal of men and women in Pakistani media texts using corpus linguistics and feminist critical discourse analysis. This research study analyzed a corpus of English newspapers having three million words in length and found strong gender differences in terms of frequency, collocations, and semantic prosodies.

Men were often overrepresented and associated with leadership and being courageous and rational. Women on their part were underrepresented and portrayed as objects that needed to be rescued by men due to their inability to act on their own. This negative connotation was evident since women were perceived more as victims rather than as agents. However, despite the differences in media houses, the prevailing trend was patriarchal.

Such findings confirm the ideological significance of the media role in sustaining gender stereotypes. The findings also emphasize the usefulness of corpus-based discourse analysis for revealing hidden linguistic prejudices that create macro-level structures of inequality.

The study brings insights into the gender prejudice found in the case of the Pakistani media with regards to the English language, which has been ignored. It illustrates the efficiency of combining corpus linguistics with feminist CDA in one approach. It contributes to global debates about media and gender through providing localized information about South Asia.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Future Recommendations

This study proves that there is a need for a critical reflection on the discourse of media in Pakistan. This study makes a contribution to the field of linguistics and, on a broader level, to gender equality activism by exposing the manner in which gender inequalities are subtly embedded in language. Some examples of cross-linguistic analysis that can be undertaken for this study include analyzing the Urdu newspapers, TV stations, and social media in contrast with English media in terms of how they depict gender issues. In addition, the use of a multimodal approach can enable the use of linguistic and visual features for diachronic corpora analysis.

References

Agha, S. (2020). Digital media and gender politics in South Asia: A discourse analysis. *Journal of Media Studies, 35*(2), 45–61.
Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.
Baker, P. (2014). *Using corpora to analyze gender*. London: Bloomsbury.
Bonyadi, A., & Samuel, M. (2013). Headlines in newspaper editorials: A contrastive study. *SAGE Open, 3*(2), 1–10.
Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. Routledge.
Cameron, D. (2007). *The myth of Mars and Venus*. Oxford University Press.
Evert, S., & Hardie, A. (2011). Twenty-first century corpus linguistics: Prospects and challenges. *International Journal of Corpus Linguistics, 16*(2), 139–166.
Fairclough, N. (1995). *Media discourse*. London: Edward Arnold.
Gill, R. (2007). *Gender and the media*. Polity Press.
Jaworska, S., & Krishnamurthy, R. (2012). On the F-word: A corpus-based analysis of the media representation of feminism. *Discourse & Society, 23*(4), 401–431.
Khursheed, S., & Kamran, S. (2019). Women in Pakistani television dramas: A critical discourse analysis. *Journal of Gender Studies, 28*(6), 707–722.
Lazar, M. M. (2005). Feminist critical discourse analysis: Gender, power and ideology in discourse. Palgrave Macmillan.
Mills, S. (2008). *Language and sexism*. Cambridge University Press.
O'Halloran, K. (2010). How to use corpus linguistics in the study of media discourse. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 563–577). Routledge.
Pearce, M. (2008). Investigating the collocational behaviour of MAN and WOMAN in the BNC using Sketch Engine. *Corpora, 3*(1), 1–29.
Roohani, A., & Barjesteh, H. (2018). Gender representation in EFL textbooks: A corpus-based approach. *Journal of Language and Gender, 12*(1), 1–22.
Shafiq, H., & Gillani, S. (2021). Gender stereotyping in Pakistani print media: A corpus-based study. *Pakistan Journal of Language and Society, 9*(2), 15–32.
Sunderland, J. (2010). *Language, gender and children's fiction*. Continuum.
Tuchman, G. (1978). The symbolic annihilation of women by the mass media. In G. Tuchman et al. (Eds.), *Hearth and home: Images of women in the mass media* (pp. 3–38). Oxford University Press.
Van Dijk, T. A. (2008). *Discourse and power*. Palgrave Macmillan.
Additional references from corpus studies on South Asia (e.g., regional media discourse) may be added to strengthen coverage.