



Gender-Based Violence in Pakistani Print Media: A Corpus Linguistics Analysis

Dilawaiz Bashir¹, Amna Hassan², Muhammad Nadeem Anwar³

¹Department of Linguistics and Communications (DLC), University of Management and Technology (UMT), Sialkot, Pakistan

²Department of Linguistics and Communications (DLC), University of Management and Technology (UMT), Sialkot, Pakistan

³University of Management and Technology (UMT), Sialkot, Pakistan

ARTICLE INFO

Keywords:

Gender-based violence, print media, Pakistani English newspapers, corpus

Corresponding Author: Muhammad Nadeem Anwar

University of Management
and Technology (UMT),
Sialkot, Pakistan

ABSTRACT

This paper seeks to explore how Gender-Based Violence (GBV) is portrayed in the Pakistani print media. The study's main goal is to investigate how linguistic choices are selected while covering GBV in the news stories of Pakistani English newspapers. The methodology used by the researchers is mixed-methods. The corpus of news stories on the said topics was collected from Pakistani English newspapers between 2021 and 2025 and analyzed through Sketch Engine. The review was based on the keywords such as rape, murder, victim, and transgender. The most important findings indicate that although newspapers report on extreme cases of GBV, they mostly use neutral and remote words. It is also biased, often neglecting important structural problems, including child marriage, activism, and transgender violence. This individualistic approach to criminal acts in lieu of macro patterns in the society implies that more considerate reporting is necessary. The paper ends by recommending the media practices that employ delicate words that are not only informative but also create more awareness and lead to societal transformation.

Introduction

Gender-Based Violence (GBV) has become one of the most significant social issues in Pakistan, and it comprises domestic violence, honor killings, sexual harassment, acid attacks, and child marriage (Ahmed, Ullah, and Noreen 2024). Although the legal changes and campaigns have been made, the media discourses of GBV reveal strong self-beliefs of patriarchy. A popular perception is created with the language of the print press, which supports or demolishes existing

gender ideologies (Asghar and Akhter 2022). The media rhetoric can shape the perception of violence in the society by either perpetuating the notion that people must blame the victim or evoking compassion and action.

The print media is a potent source in Pakistan, which constructs societal problems and social issues in terms of language, including word frequency, collocation and syntactic structure. However, the reporting makes the gendered violations less serious or shifts the blame to the victims and, therefore, guarantees the culture of impunity and silence (Yasmeen, Bukhari, Abubakar, and Masso 2019). The high frequency of the usage of the euphemistic descriptions and the negative framing instruments are valuable to support the necessity of the organized linguistic study of this reporting style.

In order to address this gap, the current study uses corpus linguistics, but it relies on Sketch Engine in its narrowest sense to analyze how Pakistani newspapers portray GBV. Through common words, collocations, and grammar structures, the present research determines the common patterns of describing victims, perpetrators, and violence itself. It is hoped that the findings will inform the healthier determination of gender, media responsibility, and the need to frame linguistics in the Pakistani socio-cultural setting. Finally, this study attempts to show the way in which the discourse of print media reproduces or subverts gendered inequalities via language.

The current research study is an attempt to address gender-based violence (GBV) in Pakistan, emphasizing problems including rape, honour murders, and domestic abuse. It highlights how the media plays a crucial role in influencing how people view violence against women, which may either promote societal change or perpetuate prejudices. The media's depiction of women as victims is patriarchal, according to current studies, however corpus linguistics is not adequately used for further in-depth examination. By using methodical, data-driven techniques to examine linguistic depictions of GBV in Pakistani print media, the study seeks to fill this gap by identifying biases and framing patterns that shape public opinion and sustain the cycle of violence.

Research Objectives

The research objectives of the study are as follows:

- To analyze the frequency and patterns of collocation of meaningful words employed to identify victims and perpetrators of violence in the press.
- To look at the ways in which socio-cultural perceptions of gender and violence in Pakistan are reproduced and consolidated in media discourses.

Literature Review

Media has a pivotal role in the development of the general perception regarding gender-based violence (GBV), both through its coverage of the events and the use of language it uses. Recent investigations in Pakistan have also started to integrate both verbal and corpus-grounded devices to reveal how these options are indicative of larger social stations and power arrangements. Asghar and Akhter (2022) investigated unsexed verbal particulars and collocates through 270-paper analysis of The News International, Dawn, and Daily Times. This exploration established the fact that the women are always peripheral beings in news media, are under-represented, and their representation is often mediated by unsexed notions.

Ahmed et al. (2024) constituted a study of how colorful types of violence conditional upon honor killings, forced marriage, and domestic violence are linguistically assembled in Pakistani journals between 2017 and 2018. The paper found that patriarchal morals are supported by the repetition of words referring to violence and through the organization which deprives womanish

subjects of their agency. These partake in disclosing patterns of sexism and marginalization in entrenched in the media language corpus that it goes without saying like the way women are so often framed as nonresistant or deposed as victims and not actors.

Akram and Yasmin (2023) explore how Noor Mukaddam's high-profile case of sexual violence was covered in the print media. The study applied transitivity and appraisal analysis to demonstrate that certain reports have a tendency to downplay or distort certain details of the case, hence, supporting rape myths. This, though, is not strictly corpus-grounded, but it gives insight into particular patterns of verbal and grammatical reportage of violence. A different study was more focused on domestic violence, specifically News Media Representations of Domestic Violence against Women in Pakistan (Ali and Pasha). This investigation, of 87 reports in leading Urdu journals, demonstrates how ethical principles are continually violated, as also how reporting is inclined towards positioning victims in traditional gender spaces, often in ways that refer to some element of bias or inequality.

Even though these studies are widely generalizable and provide a practical theoretical or empirical basis, certain gaps have been identified, which the present exploration can fill with results from many different studies focused astronomically on the topic of women's representation, patriarchy, or gender ideas, instead of being zoomed into the different terms used in the topic of GBV (e.g., rape, hijacking, victim, murder) across a variety of journals and across the combined data sets. Even though the authors such as Asghar and Akhter (2022) and Ahmed et al. (2024) employ collocational analysis, the majority of their studies examine the contextual application of GBV words in various narratives to understand the victimization of victims and offenders through the application of words.

The majority of qualitative or mixed styles are not quantitative verbal descriptions but rather completely corpus-driven verbal conclusions of GBV. Many reports either discuss case studies or are converse analyses of fabrics, but few studies entirely examine frequencies, expression and concordance of GBV in a corpus in Pakistani print media. In this way, the work, on the basis of the small technical corpus and strapping on the GBV-related verbal and syntactic patterns (frequency, expression, concordance) of various journals, solves the very place where the exploration is the thinner element.

Even though extensive research has been conducted on gender-based violence (GBV) in Pakistan, most of its past studies have been predominantly sociological, psychological, or feminist in nature, examining the causes, outcomes, and implications on policies. Nevertheless, it is not well documented as to how GBV is portrayed in media through the use of words. Many studies in media research incorporate qualitative analysis or critical discourse as their methods; however, few studies have incorporated corpus linguistics to demonstrate the frequency of some words' occurrence, the frequency of their usage in combination with another word, and how these words relate to the remainder of the news story.

Furthermore, the current research is frequently preoccupied with the visual or narrative meaning of GBV as opposed to the wording and sentence construction that silently obeys the victim and perpetrator attitudes towards people. The fact that language patterns are not paid attention to indicates that the argument of language in influencing social attitudes towards GBV is not comprehensively understood. Due to this fact, this work seeks to fill this problem gap through the corpus linguistics applied to a given collection of Pakistani newspaper articles. Using words such as rape, murder, victim, and transgender in the media sources, the study offers an actual instance of the typical language description and the influence the selection of language has on the coverage of GBV in the Pakistani print media.

Methodology

The research has taken a corpus-based approach in determining how Gender-Based Violence (GBV) has been linguistically constructed in the Pakistani print media. It has tried to strike repetitive words, word sequences and contextual relationships in a bid to realize the impact the words used in coverage have on individuals' minds towards such events. Four of the most widely read English-language Pakistani newspapers, i.e., Dawn, The Express Tribune, The Nation, and The News, were used as sources of information. The said newspapers were used based on the fact that they have a large following, and they are a journal used at a national scale. Rape, murder, abduction, and domestic violence news headlines, which are all gender-based offences, were used. The sum of words collected was around 4,219.

The selected news stories were organized into a specialized mini-corpus with the help of Sketch Engine, a trustworthy tool of linguistic data analysis. The texts were pre-processed to eliminate the non-relevant parts of the texts, such as headlines, captions, adverts and repetition of similar information. Only the article body was kept to ensure that the study was an accurate representation of the original text written by journalists.

The corpus was quantitatively and qualitatively examined using linguistic techniques. Quantitatively, the most frequent words that appeared and the frequency with which they appeared were established, as well as their collocations, to guarantee patterns of co-occurrence. Qualitatively, the research analyzed lines of concordance to look for context use of the keywords. The keywords rape, murder, victim, trial court, and transgender were used to illustrate how GBV news is constructed in the media. Intersectional contextual and statistical analyses gave a critical review of GBV reporting linguistic strategies. There was no private or confidential data that needed to be processed because all the content used in this research was de-identified from the public domain of the newspaper archives. Analytical research was conducted in a neutral fashion.

Data Analysis

This part summarizes the findings of a linguistic examination of a collection of papers based on sampled Pakistani newspapers: Dawn, the Express Tribune, The Nation and The News. The corpus consisted of approximately 4,219 words and was processed through the use of Sketch Engine to determine common words, collocations, and patterns in the description of Gender-Based Violence (GBV) in media texts. This analysis was aimed at learning how the media language employed affects the covering and presentation of violence against women.

The terms *victim*, *rape*, *murder*, *transgender* and *abduction* were the most common words in the data set. The main themes of GBV are reflected through these terms.

Figure 1: *Frequency and collocational pattern of the term victim*

] ① doc#0 outskirts of Lahore. </s><s>Authorities are actively working to locate the **victim** and her family, with legal proceedings already underway. </s><s>The in
] ① doc#0 ig found guilty of homicide. </s><s>In their order, Italian judges said the **victim's** mother "may have been the person who actually carried out the murder
] ① doc#0 ><s>The killing was "spur of the moment" and in a fit of anger over the **victim's** desire to leave home and live with her boyfriend, the Italian agency ad
] ① doc#0 rome and live with her boyfriend, the Italian agency added. </s><s>The **victim's** uncle, who was extradited from France to face the trial, was also sente
] ① doc#0 ears, while her two cousins were acquitted. </s><s>Sara Inam murder: **Victim's** father challenges mother-in-law's acquittal in IHC: The father of Sarah I
] ① doc#0 – has challenged in the Islamabad High Court (IHC) the acquittal of the **victim's** mother-in-law Samina Shah, who was the co-accused in the case. </s>
] ① doc#0 discovered, Italian news agency Ansa reported on Monday. </s><s>The **victim**, Saman Abbas, went missing from the Italian city of Reggio Emilia on A
] ① doc#0 lded. </s><s>Four days ago, on Nov 25, police in Pakistan arrested the **victim's** father, Shabbir Abbas, from his village in Punjab on a tip-off by Italian a
] ① doc#0 rder, had left Italy after the incident. </s><s>However, in Sept 2021, the **victim's** uncle – who reportedly strangled her – was arrested in Paris, AP report
] ① doc#0 ccused cousins is currently in jail in Italy. </s><s>According to AFP, the **victim** rebelled against her family's plan to have her wedded to a cousin in Pai
] ① doc#0 Pfau Civil Hospital Karachi to fulfil medico-legal formalities. </s><s>The **victim** was mother of a four-year-old son. </s><s>Transgender anchor survive:

According to Figure 1, the most frequently used term in the corpus was the word *victim* which was used eleven times. It would be used with such family-related words like *mother*, *father*, and *uncle* and also with names of individual people like *Saman Abbas*. This tendency demonstrates that victims are most likely to appear in the media as a part of the family or social environment, which can evoke the feelings of sympathy, yet the focus is likely to turn towards the systemic problems concerning GBV.

Figure 2: Frequency and collocational pattern of the word murder

] ① doc#0 anhunt in Pakistan have appeared in court in the UK charged with her **murder**. </s><s>The court hearing came after the three were arrested on Wed
] ① doc#0 in November 2022. </s><s>Nazia Shaheen, 51, had fled Italy after the **murder**, and an international arrest warrant was out for her, Ansa added. </s><s>
] ① doc#0 3, her husband, Shabbir Abbas, who also arrived in Pakistan after the **murder**, was extradited to Italy and is now in prison. </s><s>The couple was s
] ① doc#0 ctim's mother "may have been the person who actually carried out the **murder**," according to Ansa. </s><s>The judges also ruled out that the motiva
] ① doc#0 g to Ansa. </s><s>The judges also ruled out that the motivation for the **murder** was Ms Saman's refusal to marry an older man in Pakistan, as has be
] ① doc#0 f to 14 years, while her two cousins were acquitted. </s><s>Sara Inam **murder**: Victim's father challenges mother-in-law's acquittal in IHC: The father
] ① doc#0 ittial in IHC: The father of Sarah Inam – a Canadian national who was **murdered** by her husband Shahnawaz Amir in Islamabad in Sept 2022 – has cha
] ① doc#0 / after arriving in the capital from Dubai where she worked, Sarah was **murdered** at a farmhouse located in Shahzad Town where Shahnawaz – the son
] ① doc#0 2023, an Islamabad trial court sentenced Shahnawaz to death for the **murder** of Sarah and imposed a fine of Rs1 million on him but acquitted Samir
] ① doc#0 Sara's father maintained that Samina was indeed an accomplice in the **murder** of his daughter. </s><s>He clarified that his appeal was against only th
] ① doc#0 e present in the farmhouse from 8pm to 10am when his daughter was **murdered**. </s><s>"Shahnawaz Amir did kill my daughter but his mother did assi
] ① doc#0 Ms Abbas's parents, uncle and two cousins, who were accused of her **murder**, had left Italy after the incident. </s><s>However, in Sept 2021, the vic
] ① doc#0 who asked for proof of her wrongdoing. 7 years old girl got raped and **murdered** in Karachi's quaidabad: The body of a 7-year-old girl, who was raped
] ① doc#0 arachi's quaidabad: The body of a 7-year-old girl, who was raped and **murdered**, was found in Karachi's Quaidabad area on Friday, police and hospita
] ① doc#0 wife over Rs 300 dispute in Karachi: KARACHI: A young woman was **murdered** allegedly by her husband, who reportedly pushed her from their fourth
] ① doc#0 scene. </s><s>He was later apprehended by police and charged with **murder**. </s><s>A complex case of Karachi teenage girl's alleged kidnapping

Figure 2 suggests that the term *murder* was used ten times, with the majority of the usage being in the passive tense, like *her murder was executed*. This trend indicates that the media tend to hide the identity of the offender and emphasize more on the incident than the man.

Figure 3: Frequency and collocational pattern of the word murder

① doc#0	ed investigation.	</s><s>Security guard arrested in Punjab over alleged	rape	of college student: Punjab police arrested a security guard from a privat	📄	
① doc#0	guard from a private college in the Gulberg area on Sunday, accused of	raping	a first-year student.	</s><s>The suspect, identified as Aun, fled after the	📄	
① doc#0	be Farooqi, who asked for proof of her wrongdoing. 7 years old girl got	raped	and murdered in Karachi's quaidabad	: The body of a 7-year-old girl, wt	📄	
① doc#0	ordered in Karachi's quaidabad	: The body of a 7-year-old girl, who was	raped	and murdered, was found in Karachi's Quaidabad area on Friday, police	📄	
① doc#0	Child Health (NICH).	</s><s>"The examination revealed violent vaginal	rape	and physical injuries," she had said.	</s><s>Man held for killing wife ove	📄
① doc#0	lemnisation of the marriage	</s><s>They were charged with abduction,	rape	, and forced marriage.	</s><s>May 8, 2022: The girl's father Mehdi Ali K	📄

According to Figure 3, the term *rape* was used three times, and it was combined with murder in the sentences, like *raping a first-year student* and *raping and murdering*. These combinations underline the fact that the focus of the media on the physical violence bears more emphasis, as opposed to the psychological or emotional outcomes of the physical violence.

Figure 4: *Frequency and collocational pattern of the word transgender*

① doc#0	n custody for 10 days during an investigation.	</s><s>MARDAN: Two transgender persons were stabbed to death near the police lines here, sources sa	📄
① doc#0	tackers escaped the area after the incident.	</s><s>Focal person for transgender persons' association, Mardan Anmol said that the slain persons along	📄
① doc#0	police lines when they were attacked.	</s><s>She said that the other transgender persons hid themselves in a bathroom to save their lives. </s><s>The	📄
① doc#0	lialities.	</s><s>The victim was mother of a four-year-old son. </s><s> Transgender anchor survives gun attack: LAHORE: Marvia Malik, Pakistan's first tr	📄
① doc#0	r anchor survives gun attack: LAHORE: Marvia Malik, Pakistan's first	transgender news anchor, survived a gun attack on Thursday night outside her res	📄
① doc#0	Is and messages from unknown numbers for raising her voice for the	transgender community. </s><s>Ms Malik said she left Lahore out of fear for her lif	📄

Figure 4 reflects that the word *transgender* was used six times. On the other hand, the word *abduction* was used twice, and it was often combined with some words such as *alleged*, *rape* and *elopement*. Such lexical combination implies that media coverage occasionally mixes coercion with consent, which can have an impact on the attitude of the population to such instances. Other similar terms, like *minor* and *underage* were less frequently used but were applied to girls and daughters with regard to, say, child marriage and homelessness. Their under-representation indicates that the media's attention on the vulnerable age groups is scarce and not well taken care of.

The use of legal and institutional words was also identified in the data set. The terms like *trial* and *court* were used three times, and they were mostly combined with the names of places like *Islamabad*, and it was always in the context of, e.g. *trial court of*. The phrase *FIR* was repeated four times, and the word was frequently combined with such verbs such as *file*, *register* and *cancel*, which meant that emphasis was made on the procedural side of legal practices. Other terms such as *acquittal* and *sentence* were used 3-4 times and were usually combined with names of real people and nothing about the judges and legal representatives. This alludes to a narrative where there is a tendency to report legal results without covering accountability or the rationale for it. Shelter home was mentioned four times, and most often it was framed in the prepositional forms *to*, *as well as*, or *by*, making it seem more of a place to be moved to rather than a place to receive assistance or in the form of rehabilitation.

In general, the discussion indicates that the Pakistani print media tends to adopt a neutral or passive language when the issue is of gender-based violence. The use of words such as *rape*, *murder*, and *victim* is central to the discussion, but in most instances these terms are represented without considering the aspects of agency, emotional aspects, or contextual information. The lesser usage of the terms connected to the protection and rehabilitation indicates that the media attention is paid to the event itself instead of discussing social problems and preventive interventions in general.

The analysis revealed that the words *rape*, *victim*, *murder*, *woman* and *hijacking* occurred most of the time in the dataset. The frequent use of these words means that there is a harmonious concentration on violent events and victims. And *rape* as often appeared with *victim*, *minor*, *brutal* and *incident* as it pressed the intensity and felonious architecture of similar acts. The use of the term murder often passed with the words like *woman*, *killed* and *body*, which had a tendency to report GBV in terms of death or serious physical harm. The patterns of expression indicated that *woman* was mostly combined with *killed*, *set up*, *allegedly*, and *domestic* which show the linguistic connection of women with violence and domesticity. These collaborators are indication that the coverage often depicts women as unresponsive figures who are influenced by violent conditions instead of personalities with agency.

Institutional and legal terms *police*, *FIR*, *court* and *trial* were also mentioned with high frequency. *FIR* was most commonly used together with *filed*, *registered* and *cancelled*; the words *trial court* were always used together with *judgment*. This is an indication of a procedural and legal vulnerability in the reporting of the GBV cases; the institutional conduct is given more attention than the individual or social limitations. There were very few or no term equivalences such as *survivor* or *activist* represented in the corpus, but the word *victim* was used continuously used in many cases, was also associated with home words such as *mama*, *son*, and *woman*. This implies that women are generally represented in terms of relationships, which support their individualities in domestic and family environments.

On the whole, the findings show that Pakistani print media, as a rule, resort to verbal and grammatical patterns which focus on violent acts and institutional reactions as well as domestic personalities. The language employed in reporting leans towards inflicting violent lives in a report and also providing minimal visibility to the voices of women or personalized points of view.

Conclusion

This study examined the coverage of gender-based violence in Pakistani newspapers. It collected a corpus of 4219 words from Dawn, the Express Tribune, The Nation and The News International. This study employed Sketch Engine to investigate the frequency of occurrence of words, the words that were co-occurrences, and the usage of words in sentences. The findings indicated that the news items pay more attention to the acts of violence and not the individuals themselves. There were such words as *rape*, *murder*, *abduction*, and *victim*, which indicated the emphasis on the brutality of violence and the position of a victim. Who did it was concealed by the common use of passive voice, such as when one says that the woman was raped or was found dead. The words like *police*, *FIR* (*First Information Report*), and *court* were frequently mentioned, which indicate that reporting is highly procedural.

Generally, gender-based violence is largely represented in Pakistani newspapers but in a manner that focuses on legal provisions and incidences rather than on how it affects women and how women see them. The legal words demonstrate that the choice of language may alter the attitude of the news and influence how human beings perceive these gender problems. This study wasn't perfect. The quantity of the text that it was reading was not so much, and they were just reading English newspapers, which might not reveal the entire image of what was being said in Pakistan. It also just examined words and the use of mixed-up words but did not take into consideration other features such as pictures, headlines, or the manner in which the stories were framed.

In future research, it would be logical to take a larger stack of text from both English and Urdu newspapers in order to get a clearer picture regarding the comparison of the languages. Online and social media could also be chosen as a source of information where researchers could find

how gender-based violence is represented on various resources. The interview with the journalists or editors may also provide useful insights as to why they choose to cover these stories in the manner they do. In conclusion, this study indicates that there are some major patterns of language use in the reporting of gender-based violence. It further informs us that we should conduct more comprehensive and systematic research in order to fully understand the role of language in influencing the way in which individuals perceive gender and justice within Pakistani media.

References

- Ahmed, N., A. Ullah, and S. Noreen. (2024). The Discourse of Patriarchy, Gender, and Violence in Pakistani English Print Media: A Corpus-Based Study. *International Journal of Social Science Archives (IJSSA)*, 7(2).
- Akram, N., and M. Yasmin. (2023). Media Portrayal and Perpetuation of Rape Myths: A Systemic Functional Linguistics Analysis of Noor Mukaddam's Case in Pakistani Print Media. *University of Chitral Journal of Linguistics and Literature*, 7(2), 199–213.
- Asghar, S., and U. Akhter. (2022). Representation of women: a corpus-based analysis of Pakistani english newspapers. *Journal of Communication and Cultural Trends*, vol. 4, no. 2, pp. 61–80.
- Yasmeen, T., S. Bukhari, S. Abubakar, and M. Masso. (2019). Media Representation of Gender and Power in Pakistan: A Critical Review. *Review of Law and Social Sciences*, 2(1), 77–90.