

Journal of Religion & Society (JR&S)

Available Online:

<https://islamicreligious.com/index.php/Journal/index>

Print ISSN: 3006-1296 Online ISSN: 3006-130X

Platform & Workflow by: [Open Journal Systems](#)**Violence Against Women in Pakistan's English-Language Press: A Corpus-Assisted Discourse Study of Dawn Newspaper Reporting in 2021****Muhammad Hamzah Masood**

Department of Linguistics and Communications, School of Liberal Arts, University of Management and Technology, Lahore

Abstract

This study examines how violence against women (VAW) was discursively constructed in Dawn, a leading Pakistani English-language daily, during 2021, a year marked by the murder of Noor Mukadam, a mob assault on a woman near Minar-e-Pakistan and controversy over prime ministerial remarks linking sexual violence to women's clothing. Drawing on the synergy of corpus linguistics and critical discourse analysis, a validated corpus of 361 texts (184,207 tokens) was compiled from a full year of Dawn news output and compared with the remaining 17,681 news texts (9,458,277 tokens). Keyword, social actor, dependency-parsed agency, collocation and concordance analyses were combined with a comparison of news reports and commentary. Dawn constructed VAW predominantly as criminal-justice procedure. Perpetrators were named mainly through legal labels, women through victimhood and kinship, and 60.5% of finite clauses reporting violence were passive, most without an agent. Commentary named men, patriarchy and the state as sources of violence, distanced itself from 'honour' and contested victim-blaming, yet shared the passive grammar of news reports. Class appeared as formulaic background in news and as an object of critique in commentary. The newspaper thus both reproduced and challenged dominant gender norms, largely along genre lines.

Keywords: *violence against women, corpus-assisted discourse studies, Pakistani English-language press, social actor representation, agency, honour, victim-blaming*

Introduction

Violence against women remains one of the most pervasive human rights violations in Pakistan, where the *Global Gender Gap Report 2021* placed the country 153rd of 156 economies (World Economic Forum, 2021). Because most citizens never witness such violence directly, what they know about it is filtered largely through the news. Newspapers decide which incidents become public events, whose voices are heard and how blame is apportioned between perpetrators, victims and institutions (Meyers, 1997). A substantial body of research has shown that these decisions are realised in patterned linguistic choices, from the naming of social actors to the grammatical encoding of agency, and that such patterns can normalise violence or displace responsibility onto women themselves (Henley et al., 1995; Lloyd & Ramon, 2017; Meluzzi et al., 2021).

The year 2021 offers an unusually dense window onto these processes in Pakistan. In April and again in June, the then prime minister publicly associated sexual violence with 'obscenity' and with the way women dress, remarks that drew editorial condemnation and protests from rights groups (Texts 62, 97, 131). In July, Noor Mukadam was killed at a residence in Islamabad, a case that dominated national coverage for the rest of the year (Text 124). In August, a woman was assaulted by a crowd of men

near Minar-e-Pakistan in Lahore during Independence Day celebrations (Text 192). The same year saw the organisers of the Aurat March booked on blasphemy charges (Text 69), a government recommendation to refer a domestic violence bill to the Council of Islamic Ideology (Text 109), the rejection of a bill against forced religious conversions by a parliamentary committee (Text 284) and the adoption of new anti-rape legislation (Texts 321, 360). Each of these developments is documented in the corpus analysed here, and the texts cited are listed in Appendix A.

Research on Pakistani media representations of violence against women has grown but remains largely qualitative and case-bound, as in studies of reporting on the 2020 Lahore-Sialkot motorway gang rape (Akram & Yasmin, 2024) and the Asma Aziz case (Akram et al., 2025). Close readings of this kind cannot show whether the patterns they identify are typical of a newspaper's routine output, nor how routine crime reporting relates to the opinion pages on which the same newspaper argues about gender and power. Corpus-assisted discourse studies (CADS) were developed to connect large-scale linguistic patterns with situated critical interpretation (Baker et al., 2008; Partington et al., 2013), yet no study known to the author has applied the approach to a complete year of a Pakistani national newspaper's coverage of violence against women.

The present study addresses this gap by examining how *Dawn* discursively constructed violence against women across its 2021 news output. It pursues two research questions.

RQ1. How were victims, perpetrators and acts of violence linguistically represented in *Dawn*'s 2021 reporting of violence against women, and what patterns of agency and responsibility did these representations construct?

RQ2. How did this reporting reproduce or contest dominant discourses of honour and victim-blaming, and how did class, religion and genre shape these constructions?

The first question addresses the objective of identifying the linguistic strategies through which violence is represented, and the second addresses the objectives of examining how those representations reflect or challenge prevailing norms and power relations and of attending to intersecting forms of disadvantage.

Literature Review

Language, News and Violence Against Women

Feminist media scholarship has long argued that news coverage frames violence rather than merely reporting it. Meyers (1997) showed that reporting tends to individualise violence against women and to blame women whose conduct departs from gendered expectations. Lloyd and Ramon (2017), comparing domestic violence coverage in two British dailies a decade apart, found that one newspaper adopted a respectful position while the other sexualised violence and distinguished between deserving and undeserving victims.

A related line of work focuses on grammar. Henley et al. (1995) found that news reports of violence against women were frequently passive and that passive framing led readers to attribute less harm and less responsibility to perpetrators. Frazer and Miller (2009) identified a double standard whereby passive constructions were more common in narratives of male violence against women than of female violence against men. Meluzzi et al. (2021) combined corpus analysis of Italian newspapers with a reader survey and showed that passivisation and nominalisation background perpetrators and shift

responsibility towards third entities such as jealousy. Agency is therefore a mechanism through which responsibility is distributed, not a stylistic detail.

Corpus-Assisted Discourse Studies

Critical discourse analysis has been criticised for relying on small numbers of texts selected in ways that may confirm the analyst's expectations. Baker et al. (2008) proposed a methodological synergy in which keyword, collocation and concordance analysis identify recurrent patterns across large datasets while discourse-analytic interpretation explains them in context, each strand prompting further enquiry in the other. Partington et al. (2013) characterised CADS as the pursuit of non-obvious meaning through systematic corpus comparison, and Gabrielatos and Baker (2008) traced the collocational construction of refugees and asylum seekers across a decade of British reporting. Gabrielatos (2007) further showed that the terms used to query a news database determine what a specialised corpus contains, so their relevance should be evaluated rather than assumed.

Violence Against Women in Pakistani Media Discourse

Discourse-oriented research on Pakistani newspapers has typically examined headlines, editorials or single high-profile cases through transitivity analysis or related critical frameworks. Akram and Yasmin (2024) showed how newspaper language shaped the representation of the victim and perpetrators of the motorway gang rape, and Akram et al. (2025) identified the circulation of rape myths in reporting on intimate partner violence. What remains underexplored is the relationship between the routine, high-volume crime report and the lower-volume but ideologically explicit commentary text, which can only be traced by examining a newspaper's output as a whole. A further question concerns the category of 'honour', which frames lethal violence as the defence of family reputation, and whether a newspaper reproduces it as a neutral description or marks it as contested.

Analytical Framework

Two theoretical resources guide interpretation. van Leeuwen's (2008) framework for representing social actors distinguishes between the inclusion and exclusion of actors, between their activation and passivation, and between categorisations based on function, relation or identity, so that lexical and grammatical findings can be read as choices about who is visible, who acts and how people are classified. Crenshaw's (1991) concept of intersectionality directs attention to the ways gender interacts with other axes of power to produce distinctive experiences of violence and of institutional neglect. In Pakistan these axes include socioeconomic status, religious minority identity and rural or urban location.

Methodology

Research Design and Data Source

The study adopts a corpus-assisted discourse design in which quantitative corpus procedures and qualitative concordance analysis inform one another iteratively (Baker et al., 2008). The data were drawn from a publicly released dataset of *Dawn* online articles compiled by Dastgir et al. (n.d.) and hosted on GitHub. The 2021 file contains the headline, date, author, URL, category tag and full body text of each item. After empty and malformed records and items dated outside 2021 were removed, 26,143 records remained, and removing duplicate URLs and duplicate texts reduced this number to 22,663 unique texts. Because the headlines in the dataset carry *Dawn*'s section labels, the analysis was restricted to general news content, namely the Pakistan section, the print edition section

and unlabelled *Dawn* online items. World, Sport, Business and entertainment content was excluded. The resulting population of 18,042 texts served as the sampling frame.

Corpus Construction

A specialised corpus was built through query-based retrieval (Gabrielatos, 2007). An initial query derived from the research questions was expanded iteratively by reading relevant texts it had failed to retrieve. The final query contained 28 expressions covering generic terms such as *violence against women* and *gender-based violence*, specific forms of violence such as rape, sexual assault, harassment, acid attacks, forced marriage, forced conversion, ‘honour’ killing, *karo-kari* and *swara*, sentence-level co-occurrences of a female social actor with a verb of violence, and the names of the most prominent cases of 2021. A text was retrieved if query expressions occurred at least three times in its body, or at least once in both its headline and its body. Matches in sentences describing accidents, natural disasters or militant attacks were not counted, and texts with a foreign dateline were excluded.

Two checks evaluated the query. For recall, *Dawn*’s own editorial tag ‘Violence against women’, attached to 120 texts in the sampling frame, served as an independent benchmark, and the query retrieved 101 of them (84.2%). Two of the 19 texts it missed carried foreign datelines. For precision, all 447 retrieved texts were screened manually by headline and lead, and 86 were excluded because they did not concern violence against women or girls. These included accident reports, cases in which the only victims were men or boys, reports of violence against transgender persons, which fall outside the definition adopted here and merit separate study, book reviews and general political news. The final corpus of 361 texts implies a query precision of 80.8%, and the remaining 17,681 texts formed the reference corpus. Each corpus text was classified, using datelines and reporting cues checked manually, as a news report or as commentary, a category comprising editorials, columns, opinion articles and features. Table 1 summarises the corpora.

Table 1

Composition of the Corpora

Corpus	Texts	Tokens	Mean tokens per text
VAW corpus (all texts)	361	184,207	510.3
News reports	319	148,139	464.4
Commentary	42	36,068	858.8
Reference corpus (remaining <i>Dawn</i> news texts, 2021)	17,681	9,458,277	534.9

Note. Tokens are lower-cased word forms, including hyphenated and apostrophised forms. Commentary comprises editorials, columns, opinion articles and features. The sampling frame consisted of 18,042 unique *Dawn* news texts published in 2021.

Quantitative Procedures

Four quantitative procedures were applied. First, keywords were identified by comparing the VAW corpus with the reference corpus. The log-likelihood statistic (Dunning, 1993; Rayson & Garside, 2000) served as the measure of statistical significance, and Log Ratio, the binary logarithm of the ratio of relative frequencies, served as the measure of effect size. Words were retained if they occurred at least 10 times in at least five texts, reached $LL \geq 15.13$ ($p < .0001$) and had a Log Ratio of at least 1, which means that their relative frequency was at least twice that in the reference corpus. The 100

strongest keywords were grouped into semantic categories. Second, the labels used for victims and perpetrators were counted and normalised per 10,000 tokens. Third, the grammatical voice of clauses reporting violence was examined by parsing the corpus with spaCy version 3.8 and its *en_core_web_sm* model. For 18 verb lemmas denoting violent acts, including *rape*, *kill*, *murder*, *assault*, *harass*, *torture*, *abduct* and *behead*, each verbal occurrence was classified as an active finite clause, a passive finite clause with a by-agent, a passive finite clause without an agent, or a non-finite or other construction. A random sample of 30 finite classifications was checked manually, and 28 of them (93.3%) were correct. Fourth, collocates of selected nodes were calculated within a span of five words on either side, retaining items with a co-occurrence frequency of at least five and a mutual information score of at least 3, ranked by logDice (Rychlý, 2008).

Qualitative Procedures

Quantitative results directed attention to concordance lines and full texts. All occurrences of *honour* were coded according to whether they were marked as contested through scare quotes or *so-called*. All texts containing lexis associated with the claim that clothing or ‘obscenity’ causes sexual violence (*clothes*, *clothing*, *dress*, *attire*, *vulgarity*, *obscenity*, *fahashi*, *purdah*, *robot*) were read in full, and those engaging the claim were coded as contesting it, partially accommodating it, or relaying it through attributed voices without evaluative challenge. Markers of class, religion and locality were read in context, and selected features were compared across genres using log-likelihood.

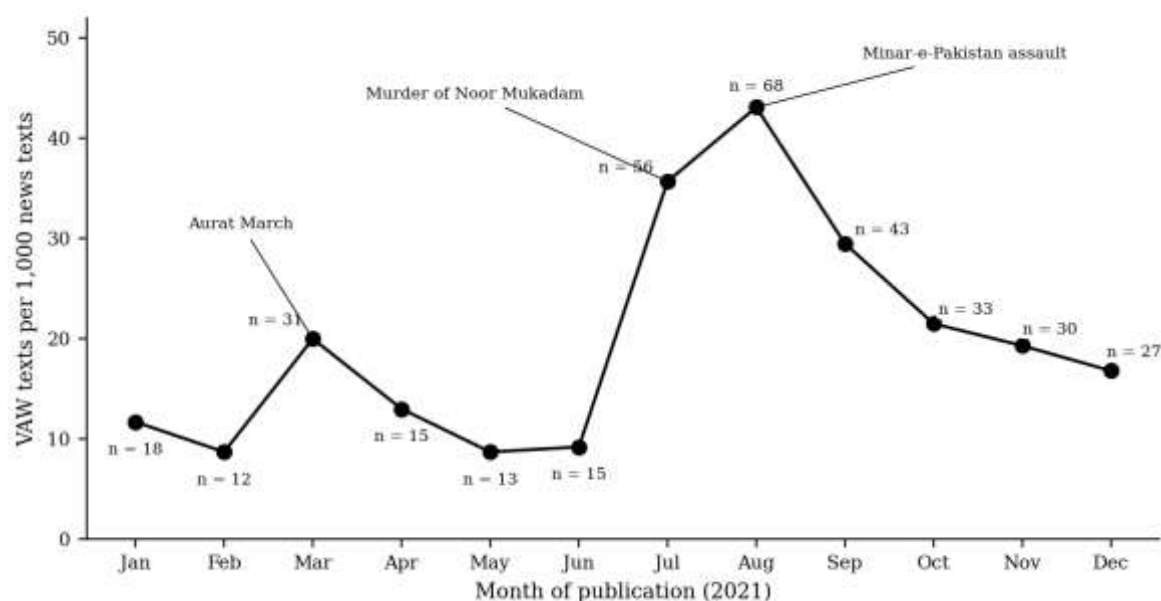
Ethical Considerations

All data are publicly available published texts. Victims are named only where their cases were already the subject of sustained national debate, no identifying detail of any child victim is reproduced, and brief quotations are identified by text number, with details in Appendix A.

Data Analysis

Distribution of Coverage Across 2021

Coverage was unevenly distributed across the year (Figure 1). It remained between 8.7 and 13.0 texts per 1,000 news texts in January, February, April, May and June, rose to 20.0 in March, when the Aurat March generated extensive reporting and counter-mobilisation, and peaked in July (35.7) and August (43.1), the months of the Noor Mukadam murder and the Minar-e-Pakistan assault. It stayed between 16.8 and 29.5 for the rest of 2021. Violence against women thus became newsworthy in bursts organised around high-profile events.

Figure 1*Monthly Distribution of Texts on Violence Against Women in Dawn, 2021*

Note. Values are VAW corpus texts per 1,000 texts in the news sampling frame for each month. Labels give raw counts of VAW texts.

Keywords

The comparison with the reference corpus yielded 583 positive keywords, and Table 2 groups the 100 strongest into six semantic categories. The most striking result is the dominance of criminal-justice procedure. Thirty-five of the 100 strongest keywords, with a combined frequency of 9,171, belong to this category, including *police* ($LL = 3,044.72$), *FIR* (first information report, $LL = 1,737.16$), *suspect* ($LL = 1,667.59$), *accused* ($LL = 1,120.12$), *remand* ($LL = 452.26$), *challan* (charge sheet, $LL = 407.32$) and *bail* ($LL = 324.10$). The hedge *allegedly* ($LL = 465.09$) also belongs here. The police, rather than women or perpetrators, are the most salient institutional actor, and violence is encountered most often as a case moving through registration, arrest, remand and trial.

Table 2*The 100 Strongest Keywords in the VAW Corpus by Semantic Category*

Category	Keywords (n)	Combined frequency	Strongest keywords (LL)
Criminal justice procedure	35	9,171	police (3,044.72), FIR (1,737.16), suspect (1,667.59), suspects (1,354.42), accused (1,120.12), case (947.74), arrested (712.44)
Women and victims as social actors	16	5,125	her (3,713.21), women (2,213.29), girl (1,257.14), woman (1,243.53), she (1,065.51), victim (983.86), daughter (904.44)
Rights, law reform and moral debate	12	2,496	Aurat (961.92), house (350.48), punishment (347.88), family (343.38), society (327.43), marriage (296.73), clothes (288.76)
Acts of violence	16	2,220	rape (2,185.68), murder (1,226.64), sexual (1,124.89), harassment (822.06), violence (798.15), raped (685.96), murdered (593.32)
Case-specific names and places	16	2,013	Zahir (2,272.45), Noor (2,223.39), Jaffer (1,739.86), Mukadam (1,242.13), Zakir (910.21), Therapyworks (623.59), Minar-i-Pakistan (225.69)
Men, kin and perpetrators	5	954	parents (732.10), men (672.89), father (532.46), husband (236.05), man (216.20)

Note. All keywords reached $LL \geq 15.13$ ($p < .0001$) and Log Ratio ≥ 1 , with a minimum frequency of 10 in at least five texts, against a reference corpus of 9,458,277 tokens. In total, 583 positive keywords met these criteria.

A second pattern is the concentration of salience in a single case. Fifteen of the strongest keywords are names and places from the Noor Mukadam case, a sixteenth refers to the Minar-e-Pakistan assault, and texts mentioning the case numbered 101, or 28.0% of the corpus. The victims of the many killings and assaults reported in brief district-level items rarely recur across texts and therefore do not surface as keywords. Keyness thus registers a hierarchy of visibility in which a few cases are narrated serially while most appear once and disappear. Men account for only five of the top 100 keywords, three of them relational terms (*parents*, *father* and *husband*).

Representation of Social Actors

Table 3 reports the principal labels used for victims and perpetrators. Women were referred to with the general terms *woman*, *women*, *girl* and *girls* 1,465 times (79.53 per 10,000 tokens), more than three times as often as men were referred to as *man* or *men* (422 times, 22.91 per 10,000). This asymmetry partly reflects the topic, but the categories through which each group was named are more revealing. Perpetrators were represented overwhelmingly through functionalised legal labels (van Leeuwen, 2008). *Suspect*, *suspects*, *suspect's* and *accused* together occurred 1,258 times (68.29 per 10,000), whereas evaluative

labels that name a person through the harm he has done, namely *culprit*, *rapist*, *killer*, *perpetrator* and *attacker* with their plural forms, occurred only 159 times (8.63 per 10,000). Legal labels are required by reporting conventions before conviction, but their near-exclusive use means that men who commit violence are constructed primarily as participants in legal procedure rather than as agents of harm.

Table 3

Selected Labels for Victims and Perpetrators in the VAW Corpus

Label group	Items	Frequency	Per 10,000 tokens
Women (general)	woman, women, girl, girls	1,465	79.53
Victimhood	victim, victims, victim's	503	27.31
Female kinship	daughter, wife, mother, sister	498	27.03
Legal role (women)	complainant	195	10.59
Survivorhood	survivor, survivors	32	1.74
Men (general)	man, men	422	22.91
Legal (perpetrators)	role suspect, suspects, suspect's, accused	1,258	68.29
Male kinship	husband, father, brother, son, cousin, in-laws	446	24.21
Evaluative (perpetrators)	culprit(s), rapist(s), killer, perpetrator(s), attacker(s)	159	8.63

Note. Kinship labels include references to relatives of victims and of accused persons as well as to perpetrators.

Women were identified predominantly through victimhood and relation. *Victim* and its forms occurred 503 times, nearly 16 times as often as *survivor* and *survivors* (32), terms that foreground endurance and agency. Female kinship terms occurred 498 times, and *daughter* alone (230) was more than twice as frequent as *wife* (107). This was largely due to a recurring formula that introduced Noor Mukadam as the daughter of a former diplomat, which appeared in 41 texts. In relational identification of this kind, a woman's newsworthiness is mediated through the status of a male relative. Where a woman was identified by something other than kinship, the category could carry evaluative weight of its own. The woman assaulted at Minar-e-Pakistan was repeatedly labelled a *TikToker* (40 occurrences in 25 texts), and, apart from her name, the strongest collocates of this label were *female* (logDice = 11.98), *companions* (11.76) and *assaulting* (11.43). Her social media activity thus became a primary identifier. One commentary text observed that responses to such violence list the many ways in which it is supposedly the woman's fault, from her clothes to her actions (Text 217), which suggests that identifiers of this kind were not neutral.

Agency and the Grammar of Violence

The parser identified 878 verbal occurrences of the 18 violent verb lemmas, distributed as shown in Table 4. Among the 428 finite clauses, 259 (60.5%) were passive and 169 (39.5%) were active. Of the passive clauses, 200 (77.2%) contained no by-agent. The most frequent grammatical subjects of passive violent verbs were *daughter* (38), *woman* (31), *Noor* (30), *girl* (19) and *she* (19). The typical clause therefore placed a woman in subject

position as the one to whom violence happened, while the man who inflicted it remained unexpressed. A related construction, *found murdered*, occurred in 38 texts and similarly foregrounds discovery over perpetration.

Table 4

Grammatical Voice of Verbal Occurrences of Violent Verbs

Construction	n	% of all occurrences	% of finite clauses
Active finite clause	169	19.2	39.5
Passive finite clause with by-agent	59	6.7	13.8
Passive finite clause without agent	200	22.8	46.7
Non-finite or other construction	450	51.3	n/a
Total	878	100.0	100.0

Note. Classification by dependency parsing with spaCy 3.8 (en_core_web_sm) for 18 verb lemmas. Manual validation of 30 randomly sampled finite classifications found 28 correct (93.3%). In texts mentioning the Noor Mukadam case, agentless passives formed 63.6% of finite clauses, compared with 36.5% in other texts.

The distribution of agentless passives departed from what might be expected. Because the Noor Mukadam case was reported in exceptional detail and the identity of the accused was never in doubt, one might anticipate that its coverage would name the perpetrator more directly. The reverse was true. In the 101 texts mentioning the case, 63.6% of finite violent-verb clauses were agentless passives, compared with 36.5% in the remaining texts. Much of this difference derives from recycled background paragraphs. A sentence quoting the first information report, in which the victim's father stated that his daughter had been brutally murdered with a sharp-edged weapon, recurred in 27 texts, and the formula reporting that she *was found murdered at a residence in Islamabad's upscale Sector F-7/4* (Text 128) recurred in 27 texts as well. Serial reporting thus fixed an agentless account of the killing as the default summary of the case, even as the same texts reported the prosecution of a named man.

Agentless and hedged constructions were also frequent in brief district reports of killings by relatives. In one report, a young woman *was shot dead allegedly for getting married against the wishes of her family* (Text 15), with the headline assigning the killing to 'family' in scare quotes. The family appears only as the source of a reason, and the killing is represented as a consequence of the woman's marital choice. The frequency of *allegedly* (189) and *alleged* (154), together 18.62 per 10,000 tokens, compounds this effect, because the hedge often attaches to the act of violence itself rather than only to the attribution of guilt.

The Discourse of Honour

The word *honour* occurred 68 times in 35 texts, and its strongest collocates were *name* ($MI = 6.08$, $\logDice = 11.91$), *killed* ($\logDice = 11.52$), *killings* (11.34) and *so-called* ($MI = 6.77$, $\logDice = 11.04$). The phraseology differed sharply by genre. In news reports, *honour* occurred 44 times, of which only 10 (22.7%) were marked as contested through scare quotes or *so-called*. The formula *in the name of honour* accounted for 14 of these occurrences, and headlines such as *Couple killed for honour in Charsadda* (Text 176) presented the category without any distancing. In commentary, by contrast, 15 of 24 occurrences (62.5%) were marked as contested. An editorial titled *Murder not honour*

argued that the state had *failed to detach the label of 'honour' from a woman's right to make her own choices* (Text 277). The contrast shows that the newspaper's editorial voice rejected the legitimating logic of the term while its routine reporting continued to reproduce it as a neutral description of motive.

Victim-Blaming and Its Contestation

Explicit metadiscourse about blame was concentrated in commentary. The expressions *victim-blaming* and *blaming the victim* occurred 24 times in 11 texts, nine of which were commentary. Sixteen texts engaged the claim that women's clothing or social 'obscurity' causes sexual violence, a claim that entered the corpus mainly through the prime minister's remarks. Eleven of these texts contested the claim. Seven were commentary texts, including an editorial that dismissed as a *myth* the idea that a woman's clothing is linked to the sexual violence against her (Text 62) and an opinion article asking how vulgarity or immodesty could trigger such crimes *and not a depraved man's mind* (Text 67). Four were news reports that gave space to critics, such as journalists, activists and opposition politicians (Texts 72, 131, 201, 203). One commentary text partially accommodated the claim, stating that *nothing that our PM said was factually wrong, just incomplete* (Text 101).

The remaining four texts were news reports that relayed the causal claim through attributed voices without evaluative challenge. One reported ruling-party legislators defending the prime minister and reproduced his statement that *if a woman is wearing very few clothes*, it would have an impact on men (Text 97). Two reported statements by religious scholars that responded to rising abuse and harassment by calling for the suppression of obscenity and nudity (Texts 207, 220). The fourth reported the prime minister's identification of obscenity as a major national problem, although its headline placed *sex crimes* in scare quotes (Text 280). The distribution indicates that contestation was voiced principally by commentary and by named critics, whereas the attributive conventions of news reporting allowed the causal claim to circulate in the newspaper's own pages without correction.

Class, Religion and the Intersections of Vulnerability

Class entered the corpus in two distinct registers. In news reports it appeared as formulaic background. The phrase *upscale Sector F-7* occurred in 38 texts, all of them news reports about the Noor Mukadam case, and the victim's father's former diplomatic status was mentioned in 41. These markers located the case within an elite urban milieu without commenting on the significance of that location. In commentary, class became an explicit object of critique. *Rich* occurred 17 times in only four texts, three of them commentary, and one column argued that the case exposed *a culture where the privileged and wealthy easily escape the law* (Text 259). Another commentary text linked the contempt directed at users of TikTok to what it termed *the intersectional curse of being a woman from a non-elite background* (Text 122), which made visible the class dimension implicit in the *TikToker* label discussed above.

Religious minority status appeared mainly in connection with forced conversion and marriage. *Forced conversion* occurred 59 times in 14 texts, and *minority* or *minorities* occurred 54 times in 17 texts, most of them news reports, which covered both parliamentary and clerical debate over the rejected anti-forced conversion bill (Text 284) and individual cases involving Christian and Hindu girls and women. In these texts, minority women appeared both in individual case reports and as the object of a contested bill. Rural and customary contexts were marked by *village* or *rural* in 46 texts and by *jirga*,

panchayat or *feudal* in 10. News reports about rural contexts that did not concern the Noor Mukadam case were short, with a median length of 296 words compared with 591 words for news reports on that case, and none of the names or places associated with them reached the top 100 keywords. The intersection of gender with elite urban status thus produced serial and extended visibility, whereas its intersection with rural location produced brief and episodic visibility.

News Reports and Commentary Compared

Table 5 compares selected features across the two genres. News reports used *police* (98.08 per 10,000 tokens) and legal labels for perpetrators (81.68) far more frequently than commentary did (16.91 and 13.31), and they relied heavily on procedural terms and the hedges *allegedly* and *alleged*. Commentary, by contrast, used *men* more than four times as often as news reports (36.87 against 8.30 per 10,000), and it used *patriarchy*, *patriarchal* and *misogyny* six times as often (12.20 against 1.89). It also referred to *society* and to the *state* at markedly higher rates. All of these differences were significant at $p < .0001$. *Victim* and its forms were also more frequent in news reports ($LL = 9.56$, $p < .01$), whereas *survivor* showed no significant difference between the genres ($LL = 0.34$).

Table 5

Selected Features in News Reports and Commentary

Feature	News f	News per 10,000	Commentary f	Commentary per 10,000	LL
Police	1,453	98.08	61	16.91	320.86
suspect(s), suspect's, accused	1,210	81.68	48	13.31	276.19
FIR, challan, remand	623	42.06	4	1.11	236.15
allegedly, alleged	334	22.55	9	2.50	91.62
victim(s), victim's	431	29.09	72	19.96	9.56
survivor(s)	27	1.82	5	1.39	0.34
Men	123	8.30	133	36.87	132.86
society, societal	106	7.16	102	28.28	90.58
patriarchy, patriarchal, misogyny, misogynistic, misogynist	28	1.89	44	12.20	59.47
State	59	3.98	59	16.36	54.55

Note. News reports comprised 148,139 tokens and commentary 36,068 tokens. $LL \geq 6.63$ corresponds to $p < .01$ and $LL \geq 15.13$ to $p < .0001$.

Grammar, however, did not follow the same division. Passive clauses accounted for 59.5% of finite violent-verb clauses in news reports and for 67.3% in commentary, and agentless passives for 46.6% and 47.3% respectively. Commentary thus named men and patriarchy as the sources of violence at the level of lexis and argument, while describing particular acts of violence in grammatical forms that were no more agentive than those of the news reports it implicitly criticised.

Discussion

The first research question asked how victims, perpetrators and acts of violence were represented and what patterns of agency and responsibility resulted. *Dawn*

constructed violence against women predominantly through a criminal-justice frame, in which police and court procedure dominated the keyword profile, perpetrators were functionalised as suspects and accused persons, and women were identified through victimhood and kinship rather than survivorhood or independent social roles. The grammatical results corroborate earlier findings that passive and agentless constructions predominate in such reporting (Frazer & Miller, 2009; Henley et al., 1995) and extend them to the Pakistani English-language press. In van Leeuwen's (2008) terms, women were passivated while men were backgrounded or excluded from the clauses describing what happened, a pattern that reduces readers' attribution of responsibility to perpetrators (Meluzzi et al., 2021). Serial reporting intensified the effect, since the most extensively covered case was summarised through recycled agentless formulae across dozens of texts. The procedural frame keeps state accountability in view and avoids sensational speculation, but it presents violence as a sequence of legal events rather than as harm done by men to women.

The second research question asked how the reporting reproduced or contested dominant discourses and how class, religion and genre shaped these constructions. Reproduction and contestation were distributed across genres. Commentary rejected the legitimating category of 'honour', named men and patriarchy as causes of violence and contested the attribution of sexual violence to women's clothing. News reports, governed by conventions of attribution and neutrality, reproduced *honour* as a description of motive in most uses and relayed the clothing claim through official and religious voices without correction. This resonates with Lloyd and Ramon's (2017) observation that one news environment can accommodate both respectful and victim-blaming positions, but it locates the division within a single newspaper rather than between titles. It also qualifies research on agency, since commentary's critique was not accompanied by more agentive grammar.

An intersectional reading adds a further dimension (Crenshaw, 1991). The elite status of the most prominent victim was encoded formulaically in news reports and turned into evidence of privileged impunity in commentary. Violence against women in rural areas, by contrast, was reported in short items that did not accumulate keyness, and violence against women from religious minorities was often subsumed within legislative controversy over forced conversion. Visibility therefore tracked social position, and the *TikTok* label shows that class could also operate against a victim.

Conclusion

This study set out to examine how violence against women was discursively constructed in *Dawn*'s 2021 reporting. Using a validated corpus of 361 texts drawn from a full year of the newspaper's news output, it found that the newspaper constructed such violence predominantly as a matter of criminal-justice procedure, in which perpetrators were functionalised as suspects, women were identified through victimhood and kinship, and violent acts were described mostly in passive clauses without an expressed agent. It also found that the newspaper both reproduced and contested dominant discourses, and that this division fell largely along genre lines. Commentary rejected the category of 'honour' and the attribution of rape to women's clothing, whereas news reports reproduced the former as a description of motive and relayed the latter through attributed voices. Class and religion shaped visibility, with elite urban cases narrated serially and the cases of rural and minority women reported episodically.

The study makes three contributions. Methodologically, it shows how an openly available news dataset can be converted into a specialised corpus whose recall and

precision are reported, and how dependency parsing can extend CADS to the analysis of agency. Empirically, it provides a year-long corpus-based account of the reporting of violence against women in a Pakistani national newspaper. Theoretically, it shows that the tension between reproducing and contesting gender norms may lie within a single newspaper, divided between the attributive conventions of news and the evaluative freedom of commentary. For journalistic practice, the findings suggest naming perpetrators in active clauses wherever facts permit, confining hedges to the attribution of guilt, marking 'honour' as contested in news as well as in commentary, following cases involving rural and minority women beyond the initial report, and supplying evidence when causal claims about sexual violence are attributed to public figures.

Several limitations should be acknowledged. The corpus represents one English-language newspaper with a predominantly urban and educated readership over one year, drawn from a third-party collection rather than an official archive. The query missed approximately one sixth of the texts *Dawn* itself tagged as concerning violence against women, voice classification was accurate in 93.3% of validated cases, and violence against transgender persons was excluded by definition. Future research should compare *Dawn* with titles such as *The Express Tribune* and *The News International*, extend the analysis to the Urdu press and trace change across several years, particularly in relation to the legislative developments of 2021.

References

- Akram, N., & Yasmin, M. (2024). Media portrayal of sexual violence in Pakistan: A critical discourse analysis of the Lahore-Sialkot motorway incident. *Women's Studies International Forum*, 106, Article 102955. <https://doi.org/10.1016/j.wsif.2024.102955>
- Akram, N., Yasmin, M., & Sohail, A. (2025). Perpetuation of rape myths through news reporting on intimate partner violence: A transitivity analysis of Asma Aziz case. *Women's Studies International Forum*, 109, Article 103038.
- Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyżanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, 19(3), 273–306. <https://doi.org/10.1177/0957926508088962>
- Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, 43(6), 1241–1299. <https://doi.org/10.2307/1229039>
- Dastgir, S., Khan, M. U., & Abbasi, A. (n.d.). *KhabarNama news ranking tool* [Dataset and source code]. GitHub. <https://github.com/SAAD-DASTGIR/KhabarNama-News-Ranking-Tool>
- Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1), 61–74.
- Frazer, A. K., & Miller, M. D. (2009). Double standards in sentence structure: Passive voice in narratives describing domestic violence. *Journal of Language and Social Psychology*, 28, 62–71. <https://doi.org/10.1177/0261927X08325883>
- Gabrielatos, C. (2007). Selecting query terms to build a specialised corpus from a restricted-access database. *ICAME Journal*, 31, 5–44.
- Gabrielatos, C., & Baker, P. (2008). Fleeing, sneaking, flooding: A corpus analysis of discursive constructions of refugees and asylum seekers in the UK press, 1996–2005. *Journal of English Linguistics*, 36(1), 5–38.
- Henley, N. M., Miller, M., & Beazley, J. A. (1995). Syntax, semantics, and sexual violence: Agency and the passive voice. *Journal of Language and Social Psychology*, 14(1–2), 60–84. <https://doi.org/10.1177/0261927X95141004>
- Lloyd, M., & Ramon, S. (2017). Smoke and mirrors: U.K. newspaper representations of intimate partner domestic violence. *Violence Against Women*, 23(1), 114–139. <https://doi.org/10.1177/1077801216634468>
- Meluzzi, C., Pinelli, E., Valvason, E., & Zanchi, C. (2021). Responsibility attribution in gender-based domestic violence: A study bridging corpus-assisted discourse analysis and readers' perception. *Journal of Pragmatics*, 185, 73–92. <https://doi.org/10.1016/j.pragma.2021.07.023>
- Meyers, M. (1997). *News coverage of violence against women: Engendering blame*. Sage.
- Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies (CADS)*. John Benjamins. <https://doi.org/10.1075/scl.55>
- Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. In *Proceedings of the Workshop on Comparing Corpora* (pp. 1–6). Association for Computational Linguistics. <https://doi.org/10.3115/1117729.1117730>

- Rychlý, P. (2008). A lexicographer-friendly association score. In P. Sojka & A. Horák (Eds.), *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2008* (pp. 6–9). Masaryk University.
- van Leeuwen, T. (2008). *Discourse and practice: New tools for critical discourse analysis*. Oxford University Press.
- World Economic Forum. (2021). *Global gender gap report 2021*.
<https://www.weforum.org/reports/global-gender-gap-report-2021>