



# A Corpus-Assisted Critical Discourse Analysis of Ideology in Pakistani and International Digital News Headlines

## Research Article

Correspondence:	Ali Raza < <a href="mailto:aliraza9924@gmail.com">aliraza9924@gmail.com</a> >	M.Phil Scholar, Department of English, University of Okara, Okara, Pakistan.
	Dr. Muhammad Ahmad < <a href="mailto:ahmad453@gmail.com">ahmad453@gmail.com</a> >	Visiting Lecturer, Department of English, University of Okara, Okara, Pakistan.
	Kainat Zulfiqar < <a href="mailto:kainatzulfiqar2021@gmail.com">kainatzulfiqar2021@gmail.com</a> >	M.Phil Scholar, Department of English, University of Okara, Okara, Pakistan.

## Publication Details

**Received:** February 25, 2026

**Accepted:** March 25, 2026

**Published:** March 31, 2026

## Abstract

This study presents a corpus-assisted critical discourse analysis of ideological bias in Pakistani and international digital news headlines to address a gap in lexicogrammatical research of Pakistani English-language media. For this purpose, this study utilizes the news headlines corpus from 12 outlets across four ideologically different sub-corpora: Pakistani (PAK), left-wing (LW), right-wing (RW), and international (INT). Therefore, this study employs a multi-method analytical protocol using AntConc and Sketch Engine, and compares the framing strategies across national and ideological contexts. Key results reveal that Pakistani headline discourse is organized around a distinctive political-institutional crisis vocabulary (accountability, crackdown, military, detained, establishment) with no direct equivalent in international sub-corpora. Semantic prosody analysis demonstrates that identical lexemes carry systematically different prosodic orientations across national contexts: accountability is positively prosodic in LW (civic virtue) but negatively prosodic in PAK (instrument of political persecution). Transitivity analysis shows that the PAK sub-corpus exhibits high active voice rates (72.3%) that encode stances toward state action while simultaneously suppress civil society (78.6% suppression rate). On the basis of these results, this



study concludes that semantic equivalence cannot be assumed across national corpora and that ideological framing operates through nationally specific lexical, collocational, and grammatical mechanisms. The implications extend to cross-national media discourse methodology, media literacy education in Pakistan, and understanding how grammatical choices reproduce political power asymmetries in Global South media systems.

**Keywords:** ideological positioning; media framing; news headlines; semantic prosody; transitivity

## 1. Introduction

The relationship between language, power, and ideology in news discourse has long occupied a central place in communication scholarship (Fairclough, 1995; Fowler, 1991; van Dijk, 1988, 2009). This relationship assumes particular analytical and civic significance in the context of Pakistan, that is, a democratic state marked by persistent political instability, a fragmented media landscape, and a journalism culture that operates under distinctive institutional restrictions including regulations by the Pakistan Electronic Media Regulatory Authority (PEMRA), pressures from political and military establishments, and economic liabilities that shape editorial independence. As Sabahat, Iqbal, Gulzar, and Arshad (2025) observe that the Pakistani media environment is characterised by intense biased alignment in which media framing in Pakistan is ideologically biased, aligns with political interests and where audience trust is strongly mediated by political affiliation. Their quantitative survey of 400 respondents demonstrated a significant positive correlation between political affiliation and media trust ( $r=0.68$ ,  $p<0.001$ ) which confirmed that ideological alignment fundamentally structures media reaction in Pakistan. These conditions make Pakistan an essential context for empirical discourse analysis. Therefore, this study comparatively explores media language particularly focusing the news headline.

The news headline is not merely a typographic convenience. It constitutes, as Dor (2003) argues, a 'relevance optimizer', that is, a maximally compressed semiotic device which selects, foregrounds, and positions the events it introduces by determining for many readers not simply whether a story will be read but within what cognitive and moral frame it will be processed. The linguistic genre of the headline is governed by its own register conventions: article deletion, the historic present, heavy nominalisation, and syntactic compression (Bednarek & Caple, 2017; Conboy, 2007). These conventions, far from being ideologically neutral, increase rather than decrease the framing effects that critical discourse analysis (CDA) has consistently identified in journalism (Caldas-Coulthard & Coulthard, 1996; Richardson, 2007). Therefore, the headline is the item in which editorial ideology is most economically and most powerfully employed.

Studies on Pakistani English newspaper language have established a distinctive register profile for this variety. Ahmad, Mahmood, Mahmood and Siddique (2019) and Ahmad, Mahmood and Farukh (2020) demonstrate through corpus-based analysis that Pakistani English newspaper editorials use modal verbs as stance markers that reflect the ideological orientations of the Pakistani press. Fatima, Siddique, Ahmad and Mahmood (2023) extend this profile through multidimensional analysis of linguistic variation in Pakistani English newspaper editorials, and observed that the register is more involved, more evaluatively explicit, and less abstractly informational than comparable British broadsheet material (i.e., the features that bear directly on the construction of ideological frames at the lexicogrammatical level). Similarly, Akhter and Shah

(2024) find how corpus-assisted analysis can reveal competing ideological frames in the representation of women rights within Pakistani English newspapers, and emphasize the value of corpus methods for studying media ideologies. Collectively, these studies establish that Pakistani news language is a distinct register with its own ideological marks, and deserves a corpus-based study.

The theoretical framework of the present study is based on three traditions: framing theory, CDA, and corpus linguistics integrated within the corpus-assisted discourse studies (CADS) framework. Framing theory, originating in Goffman's (1974) sociological account of interpretive schemata and operationalised for media analysis by Entman (1993), proposes that news texts promote particular 'problem definitions, causal interpretations, moral evaluations, and treatment recommendations' through the selective salience of linguistic and visual elements. Entman (2007) extends this account to demonstrate that 'framing bias' in media systems (i.e., the systematic favouring of particular definitional frames) constitutes a form of 'distributional power' that shapes the cognitive environment within which citizens form political judgements. In Pakistan, where Sabahat et al. (2025) observe strong correlations between media framing style and public perception of political crises ( $\beta=0.36$ ,  $p<0.001$ ), the political stakes of framing analysis are particularly important.

CDA provides the interpretive vocabulary for connecting the linguistic patterns identified by corpus analysis to their ideological and social functions. Fairclough's (1995) three-dimensional model, which conceptualises discourse simultaneously as text, discursive practice, and social practice has proven the most productive CDA framework for media discourse research (Almohawes, 2026; Khalifah, Altakhaineh & Za'rour, 2026) because it facilitates analysis at the micro-level of individual lexical choice, the meso-level of editorial and institutional production, and the macro-level of social power and ideology. Khalifah et al. (2026), in their bilingual CDA of Al Jazeera's coverage of the Gaza-Israel war, observe that a single transnational media institution can construct systematically different ideological narratives for Arabic and English-language audiences with the Arabic service employing religious terminology (e.g., 'martyrs' in 89% of articles) and active constructions foregrounding Israeli agency, while the English service used strategic passivisation and source diversity to satisfy international objectivity norms. Van Dijk's (1998) account of the ideological square (the systematic linguistic strategies of positive in-group representation and negative out-group representation) provides a specific analytical tool for the headline-level analysis pursued in this study. Halliday's (1994) systemic functional grammar (particularly its account of transitivity as a system for distributing agency across participants in material processes) supplies the grammatical framework for the agency analysis reported in Section 3.

Corpus linguistics contributes the methodological infrastructure that transforms subjective CDA judgments about individual texts into empirically grounded claims about systemic discourse trends. Sinclair's (1991) idiom principle (the notion that language forms patterned co-occurrence networks where statistical regularities constitute a primary locus of meaning) supports the collocational and semantic prosody analyses at the core of this study. McEnery and Hardie (2012) argue that the shift from individual text to corpus constitutes a 'fundamentally different relationship with linguistic evidence', one that limits analyst subjectivity and supports principled generalisation. Baker, Gabrielatos, KhosraviNik, Krzyżanowski, McEnery and Wodak (2008) operationalised this relationship in what has become the paradigm study of CADS methodology, and demonstrated that the combination of log-likelihood keyness analysis, MI-based collocate extraction, and

concordance-based close reading can reveal the systematic discursive construction of social groups (refugees and asylum seekers in their case) with a rigour that pure qualitative CDA cannot achieve. This methodological framework has been extended to national policy discourse (Almohawes, 2026), conflict reporting (Alluhaidah, 2023; Khalifah et al., 2026), and political crisis representation (Sabahat et al., 2025) which establishes CADS as the leading approach to critical empirical analysis of media discourse.

Research on Pakistani media through corpus and discourse methods remains comparatively ignored in relation to the Pakistani journalism as a sociopolitical phenomenon. While Ahmad et al. (2019, 2020), Fatima et al., (2023), and Siddique, Mahmood, and Ahmad (2023) have studied Pakistani English press language, no published study has conducted a systematic CADS investigation of Pakistani news headline discourse in comparison with international media sub-corpora. Existing work on Pakistani media framing, such as Sabahat et al. (2025), has employed mixed-methods approaches that combine limited textual analysis with quantitative audience surveys without the capacity for large-scale lexicogrammatical analysis that corpus tools afford. This gap is significant, and means that the most linguistically precise and methodologically transparent form of media bias research has not yet been applied to one of most important media systems in South Asia. This present study addresses this gap.

As Baker, Gabrielatos and McEnery (2013) demonstrate that ideological patterns in media language are most productively identified not through the analysis of a single outlet but through systematic comparison across ideologically differentiated corpora. By constructing a purpose-built corpus that compares Pakistani English-language newspaper headlines with international sub-corpora representing left-wing (LW), right-wing (RW), and editorially neutral international (INT) orientations, this study enables both internal analysis of Pakistani media discourse and its contextualisation within a broader international comparative frame. This design allows the study to answer not only "what is distinctive about Pakistani news headline language?" but "how does that distinctiveness relate to the ideological patterns characteristic of contrasting international media systems?", that is, a question with significant implications for understanding the relationship between national political context and news discourse practices.

Recent cross-national discourse research has documented the significant ways in which political and institutional context shapes the specific framing resources used in news language. Almohawes (2026), investigating the NOW Corpus, shows that international media coverage of Saudi Vision 2030 is organised around a consistent set of collocates (i.e., goals, objectives, diversification) that reflect the strategic framing ambitions of the Saudi government and its management of international discourse. Her analysis reveals that the phrase 'Saudi Vision 2030' increased from 43 mentions in 2016 to 308 in 2024 with strong lexical associations such as 'goals' (MI 5.25), 'objectives' (MI 6.50), and 'diversification' (MI 6.51) which reflect a discourse focused on strategic reform and economic transformation. These studies establish that the national and institutional contexts of news production leave systematic traces in the statistical regularities of news language that corpus methods can study. This present study contributes to this growing body of cross-national CADS research by providing the first large-scale corpus analysis of Pakistani news headline discourse within an international comparative framework. In this regard, this study addresses following research questions mapped onto the analytical dimension of the CADS framework.

1. What statistically significant lexical differences characterise the headline language of Pakistani news outlets in comparison with left-wing, right-wing, and international sub-corpora, and what dominant ideological orientations do these differences reflect?
2. How do the collocational environments and semantic prosody of lexemes differ across Pakistani and international news sub-corpora, and what framing functions do these different patterns perform in constructing distinct versions of political reality?
3. To what extent do patterns of transitivity, grammatical agency attribution, and evaluative lexis in Pakistani news headlines encode an ideologically distinctive positioning in relation to political actors and institutions as compared with the framing strategies reported in international media sub-corpora?

## **2. Materials and Methods**

### **2.1 Research Design**

This study employs a mixed-methods research design within the CADS framework (Baker et al., 2008; Partington, Duguid & Taylor, 2013) combining quantitative corpus-linguistic analysis with qualitative critical discourse interpretation. The epistemological rationale for this integration is articulated by Baker et al. (2008), who argue that CADS produces analyses that are more explicit, accountable and reproducible than either quantitative or qualitative methods, because quantitative frequency patterns limit the selection of interpretive focus while qualitative concordance reading provides the contextual depth necessary to explain those patterns. This present study applies this logic through four sequential analytical stages (i.e., frequency profiling, keyness analysis, collocate extraction, and concordance-based semantic prosody analysis) each informing and being informed by the others.

The theoretical framework integrates three components: (1) Entman's (1993, 2007) framing theory which provides the macro-level conceptual vocabulary for understanding why lexical choices in headlines matter as they determine the interpretive frame within which readers process events, assigning causal responsibility, moral evaluation, and preferred treatment recommendations; (2) Fairclough's (1995) three-dimensional CDA model, operationalised through Van Dijk's (1998) ideological square and Halliday's (1994) systemic functional grammar, which provides the analytical tools for connecting statistical linguistic patterns to social power and ideology; and (3) corpus linguistics, grounded in Sinclair's (1991) idiom principle and Brezina's (2018) statistical framework which provides the methodological framework for extracting, quantifying, and comparing the lexical patterns in which ideological meaning is embedded.

### **2.2 Corpus Development**

The news headlines comparative corpus 2023 (NHCC-2023, Table 1) was constructed according to three core corpus design principles: representativeness, comparability, and analytical adequacy (Biber, Conrad & Reppen, 1998; McEnery & Hardie, 2012). The corpus comprises four ideologically differentiated sub-corpora each consisting of 5,000 headlines drawn from three news outlets representing a consistent editorial orientation: (1) a Pakistani sub-corpus (PAK) drawn from the principal English-language dailies Dawn, The News International, and The Express Tribune; (2) a LW international sub-corpus drawn from The Guardian, The Independent, and HuffPost UK; (3) a RW international sub-corpus drawn from The Daily Telegraph, The Times, and The Daily

Mail; and (4) an International/neutral sub-corpus (INT) drawn from Reuters, BBC World Service, and Al Jazeera English. Table 1 presents the full composition statistics of the NHCC-2023.

Table 1: Corpus Distribution

Sub-Corpus	Source Outlets	Headlines	Tokens	Types	TTR	Period
<b>Pakistani</b>	Dawn; The News International; The Express Tribune	5,000	64,218	7,943	0.1236	Jan–Dec 2023
<b>Left Wing</b>	The Guardian; The Independent; HuffPost UK	5,000	67,432	8,341	0.1237	Jan–Dec 2023
<b>Right Wing</b>	The Daily Telegraph; The Times; The Daily Mail	5,000	68,215	8,612	0.1263	Jan–Dec 2023
<b>International</b>	Reuters; BBC World Service; Al Jazeera English	5,000	71,890	9,203	0.128	Jan–Dec 2023
<b>TOTAL</b>	<b>12 outlets</b>	<b>20,000</b>	<b>271,755</b>	<b>18,432</b>	<b>0.0679</b>	<b>Full year</b>

\*TTR = Type-Token Ratio

As shown in Table 1, the four sub-corpora are closely matched in headline frequency (5,000 each) and comparable in token counts. The PAK sub-corpus consists a slightly lower token count (64,218) than the international sub-corpora which reflects the characteristic compression of Pakistani English headline style reported by Fatima et al. (2023) and the telegraphic register features identified by Ahmad et al. (2019, 2020). All sub-corpora achieve Type-Token Ratios broadly comparable to those reported in previous headline corpus studies (Bednarek, 2016; Brookes & McEnery, 2019), and confirm adequate lexical diversity for statistical analysis. The combined corpus of 271,755 tokens across 20,000 headlines exceeds the analytical adequacy threshold for reliable keyness and collocate analysis at all frequency thresholds employed in the study (minimum frequency: 5; keyness LL threshold: 10.83,  $p < 0.001$ ).

The selection of Pakistani outlets reflects the three dominant voices in English-language Pakistani journalism: Dawn, Pakistan's oldest and prestigious English daily associated with a liberal, secular editorial tradition; The News International, the largest-circulation English daily adopting a centrist but politically engaged position; and The Express Tribune launched in partnership with the International Herald Tribune and associated with a broadly reformist orientation. This selection captures the mainstream of Pakistani English-language journalism without claiming to represent the full spectrum of Pakistani media (including Urdu-language press, broadcast, and digital-only outlets) with a limitation acknowledged in Section 4.3.

### 2.3 Data Collection Protocol

Headlines were collected using a systematic, automated web-scraping procedure implemented in Python 3.11 by using the requests and BeautifulSoup libraries for HTML parsing and the Scrapy framework for structured crawling of outlet RSS feeds. Collection was scheduled at 06:00, 12:00, and 18:00 UTC daily for each outlet throughout the calendar year 2023. This temporal coverage

captures the diurnal news cycle for all time zones represented in the corpus, and minimises systematic temporal bias. Each collection run retrieved headlines from the main page and all primary topical section pages (home, national, international, politics, business, and opinion-excluded). Metadata collected per headline included: *headline\_text*, *outlet*, *sub\_corpus*, *section\_tag*, *date*, *time\_of\_collection*, *url*, and *word\_count*. Opinion column headlines were systematically excluded through keyword-based section tag filtering, to ensure that the corpus represents factual news reporting rather than overtly marked opinion. Headlines below four words in length were excluded as insufficiently representative of substantive framing content. Daily automated deduplication removed identical headlines across collection runs using URL normalisation and Levenshtein distance thresholding (threshold:  $\leq 2$  character edits).

## **2.4 Corpus Pre-Processing and Development**

The collected newspaper headlines were systematically pre-processed to ensure the accuracy, consistency, and reliability of the corpus before conducting corpus-assisted discourse analysis. Initially, all headlines were converted into plain-text (.txt) files encoded in UTF-8 format to maintain compatibility across corpus analysis software. During this stage, unnecessary formatting, duplicated spaces, HTML symbols, and non-linguistic characters were removed while preserving the original wording and punctuation of the headlines.

The cleaned corpus was then subjected to word tokenisation, in which the text was segmented into individual lexical units (tokens). Tokenisation was performed using AntConc (Version 4.2.4), which automatically separates words, punctuation marks, and symbols to facilitate subsequent lexical analyses, including frequency lists, keyword analysis, collocation, concordance, and n-gram extraction (Anthony, 2022).

Following tokenisation, the corpus was annotated using the CLAWS Part-of-Speech (POS) Tagger, which assigns grammatical categories to each lexical item according to the Penn Treebank tagset. POS tagging enabled the identification of nouns, verbs, adjectives, adverbs, pronouns, and other grammatical categories required for the examination of lexical choices, transitivity patterns, agency, and evaluative language within the newspaper headlines. After the automatic tagging process, the tagged files were manually examined to identify and correct potential tagging errors involving proper nouns, abbreviations, and context-specific political terminology commonly occurring in Pakistani newspaper discourse.

The processed corpus was subsequently imported into AntConc for lexical analyses, including word frequency, keyword extraction, collocation analysis, concordance analysis, and cluster generation. In addition, the corpus was uploaded to Sketch Engine to generate word sketches, collocational networks, keyword statistics, and distributional patterns that complemented the analyses conducted in AntConc. The combined use of these corpus analysis tools enhanced the reliability of lexical and phraseological investigations by enabling cross-validation of the findings.

To examine agency representation, the tagged corpus was manually analysed using the POS-tagged output, concordance lines, and contextual examination of headlines. Each headline was coded for active and passive voice, agent foregrounding, agent suppression, and transitivity patterns based on the analytical framework adopted for this study. To ensure the reliability of the coding process, a

randomly selected sample of headlines from each sub-corpus was independently coded and compared, resulting in substantial inter-rater agreement as measured by Cohen's kappa ( $\kappa = 0.82$ ). This level of agreement indicates a high degree of consistency in the annotation and coding procedures, thereby strengthening the validity of the corpus-based analysis.

## **2.5 Corpus Analysis Tools**

### **2.5.1 AntConc**

AntConc (Version 4.2.4, Anthony, 2022) was used for frequency profiling, keyword analysis, collocate extraction, and concordance generation. The Keyword List tool was configured to use the log-likelihood (LL) statistic with a minimum significance threshold of  $LL \geq 10.83$  ( $p < 0.001$ , two-tailed) and a minimum target corpus frequency of 5 as recommended by Brezina (2018) and Gabrielatos (2018). For cross-sub-corpus keyword comparison, each sub-corpus was analysed against the pooled remaining three sub-corpora as reference, that is, a design that maximises sensitivity to sub-corpus-specific lexical distinctiveness. The Collocates tool was configured with a span of L5/R5 and  $MI \geq 3.0$  /  $t \geq 2.0$  thresholds (Stubbs, 2001).

### **2.5.2 Sketch Engine**

Sketch Engine (Kilgarriff et al., 2014) was employed primarily for its Word Sketch function which provides grammatically structured collocate profiles by distinguishing between modifiers of the target lemma, verbs for which the target is object or subject, and nominal co-occurrences. This syntactic specificity goes beyond AntConc's window-based collocate extraction to reveal the grammatical positions in which ideologically loaded associations are manifested which are essential for the transitivity and agency analyses addressed in RQ3. Church and Hanks's (1990) Mutual Information (MI) metric was used for collocate scoring with BIC adjustment and a minimum frequency threshold of 5.

## **2.6 Statistical Analysis**

All statistical thresholds were set a priori and held constant across all analyses in line with recommendations in Brezina (2018) and Gabrielatos (2018). The log-likelihood (LL) ratio test was used for keyness analysis, with  $LL \geq 10.83$  ( $p < 0.001$ ) as the significance threshold and  $LL \geq 15.13$  ( $p < 0.0001$ ) as the high-significance threshold reported in Table 2. Effect size was calculated using the %DIFF metric (Gabrielatos, 2018) which provides an interpretable estimate of the magnitude of over- or under-representation independent of corpus size.  $MI \geq 3.0$  and  $t \geq 2.0$  were applied as conjoint inclusion criteria for collocate analysis following Stubbs (2001) and Kilgarriff et al. (2014). For the semantic prosody analysis, a minimum of 50 concordance lines per item per sub-corpus were examined with each line coded for positive, negative, or neutral evaluative orientation using the criteria specified in Hunston and Thompson (2000). Inter-coder reliability for semantic prosody classification was assessed on a 20% sample ( $\kappa = 0.79$ ) which indicates substantial agreement). The transitivity annotation achieved  $\kappa = 0.82$  as noted above.

### 3. Results and Discussion

Results are presented and discussed in relation to each research question with quantitative corpus findings interpreted through the theoretical frameworks of Entman (1993, 2007), Fairclough (1995), Halliday (1994), and Van Dijk (1998). In keeping with the CADS principle of iterative movement between statistical pattern and qualitative interpretation (Baker, 2006; Partington et al., 2013), each results subsection presents corpus data and situates those data immediately within their theoretical and contextual significance.

#### 3.1 Keyword Analysis: Lexical Markers of Ideological Orientation

Keyword analysis comparing each sub-corpus against the pooled remaining three sub-corpora provided robust and theoretically interpretable keyword profiles for all four sub-corpora. Table 2 presents the 15 highest-ranking keywords for the PAK sub-corpus in comparison with international sub-corpora by reporting raw frequencies, LL scores, p-values, directionality, and the dominant semantic frames identified through concordance reading.

Table 2: Top 15 Keywords by Log-Likelihood Score

#	Keyword	Freq PAK	Freq INT	LL Score	p-value	Direction	PAK Semantic Frame	INT Semantic Frame
1	accountability	456	78 (INT)	278.4	<0.001	PAK+	State corruption/judicial crisis frame	Anti- accountability /absent
2	crackdown	389	45 (INT)	298.1	<0.001	PAK+	State repression of dissent	Legal enforcement frame
3	military	534	124 (INT)	267.3	<0.001	PAK+	Institutional power dominance	Defense/secu rity frame
4	detained	289	38 (INT)	231.6	<0.001	PAK+	Political imprisonment frame	Lawful custody frame
5	establishment	234	19 (INT)	219.4	<0.001	PAK+	Structural power critique	Absent
6	judiciary	267	43 (INT)	213.7	<0.001	PAK+	Constitutional legitimacy	Legal procedure
7	inflation	423	156 (INT)	198.6	<0.001	PAK+	Economic suffering frame	Economic metric
8	IMF	298	87 (INT)	176.3	<0.001	PAK+	External dependency frame	Bailout/refor m frame
9	rupee	198	23 (INT)	178.9	<0.001	PAK+	Currency vulnerability	Foreign exchange metric
10	terrorism	312	89 (INT)	187.4	<0.001	PAK+	Existential security threat	Regional security concern
11	austerity	34	342 (LW)	218.4	<0.001	LW+	Economic burden (peripheral)	Social justice frame
12	illegal	43	389 (RW)	287.6	<0.001	RW+	Legal status (peripheral)	Criminalisati on frame

13	climate	67	612 (LW)	312.4	<0.001	LW+	Environmental concern (limited)	Policy priority
14	sovereignty	39	287 (RW)	211.3	<0.001	RW+	National identity (rhetorical)	Constitutional frame
15	welfare	31	421 (LW)	187.6	<0.001	LW+	Social protection (limited)	State responsibility

As Table 2 shows, the keyword data reveal a Pakistani headline lexicon organised around an exceptionally prominent discourse of institutional crisis and political competition. The five highest-ranking PAK+ keywords, that is, accountability (LL = 278.4), crackdown (LL = 298.1), military (LL = 267.3), detained (LL = 231.6), and establishment (LL = 219.4) form a tightly clustered semantic domain centred on the competition between political and military institutional power, the criminalisation of political opposition, and the vulnerability of civil liberties. This cluster is not an object of a single news event but, as concordance analysis confirms, indicates the sustained coverage throughout 2023 of the Pakistani political crisis precipitated by the arrest of former Prime Minister Imran Khan in May 2023 and the subsequent crackdown on his Pakistan Tehreek-e-Insaf (PTI) political movement.

The presence of establishment (LL = 219.4) as a high-ranking PAK+ keyword deserves particular attention. This lexeme carries in Pakistani political discourse a highly specific ideological referent, that is, the informal group of military, intelligence, and security that exerts structural power over civilian governance, and that has no direct equivalent in the British press vocabulary represented by the LW and RW sub-corpora. Its frequency (234 occurrences in the PAK sub-corpus) and its near-absence in international sub-corpora (19 occurrences combined) confirms that Pakistani headline discourse operates within a political semantic field that is incommensurable with those of its international counterparts, and which encodes assumptions about the nature of Pakistani state power that are taken for granted within the domestic discourse community but require active translation for international readers. This finding extends the account of national political vocabularies in Van Dijk (1998) by demonstrating that ideological keywords in one national discourse community may be entirely absent from the semantic landscape of international journalism, not because the referent is absent but because the discursive framing of that referent is culturally specific.

The economic keyword cluster in the PAK sub-corpus, that is, inflation (LL = 198.6), IMF (LL = 176.3), rupee (LL = 178.9) reflects the particular form taken by economic crisis discourse in a Global South context: where the LW sub-corpus's economic keywords (austerity, welfare, inequality) position the state as the agent responsible for distributional failures, and the RW sub-corpus's keywords (tax burden, growth, regulations) position the state as the agent of economic constraint, the PAK sub-corpus economic vocabulary positions the Pakistani state as the patient of external financial discipline (subject to IMF conditionality, exposed by a depreciating currency, and overwhelmed by inflation) rather than as an autonomous agent of economic choice. This external-dependency framing is consistent with postcolonial accounts of development discourse (Herman & Chomsky, 2021; Mullainathan & Shleifer, 2005) and extends the framing theory literature by indicating that the causal attribution of economic phenomena in news discourse differs fundamentally between Global North and Global South media systems.

The security keyword terrorism (LL = 187.4) represents a further point of Pakistani-specific lexical distinctiveness. In the LW and RW sub-corpora, terrorism appears primarily in coverage of international conflicts and is rarely applied to domestic political events. In the PAK sub-corpus, concordance analysis reveals that terrorism co-occurs frequently with domestic political actors, judicial proceedings, and security operations in Khyber Pakhtunkhwa and Balochistan that reflects Pakistan's distinctive security environment in which the boundary between political opposition, ethnic insurgency, and religiously motivated violence is actively competed in headline discourse. This finding aligns with Khalifah et al.'s (2026) finding that conflict reporting requires analysts to attend to the specific political and cultural contexts that determine which actors are categorised as 'terrorists', 'militants', or 'resistance' which is a categorisation that, as their comparative data on Al Jazeera Arabic and English show, is never ideologically free.

### 3.2 Collocational Environments and Semantic Prosody

Table 3 presents the top ten collocates of the lemma 'government' across the four sub-corpora extracted using Word Sketch function through Sketch Engine (BIC-adjusted MI, window L5/R5, minimum frequency 5). 'Government' was selected as the analytical node because its collocational environment most directly indexes the ideological positioning of each sub-corpus in relation to institutional political authority, the central contested domain in Pakistani political discourse and a critical dimension of the LW/RW divergence documented in the international sub-corpora.

Table 3: Top 10 Collocates of the Lemma 'Government' across the Corpora

#	PAK Collocate	MI	Freq.	LW Collocate	MI	Freq.	RW Collocate	MI	Freq.
1	crackdown	8.76	267	cuts	7.89	421	growth	8.43	623
2	unconstitutional	8.43	189	austerity	7.12	342	reform	7.67	389
3	failed	7.89	312	welfare	6.98	398	policy	7.43	534
4	collapse	7.54	198	inequality	6.87	287	freedom	7.12	312
5	detained	7.12	245	spending	6.54	467	initiative	6.98	287
6	accountability	6.87	389	workers	6.32	534	sovereignty	6.76	234
7	illegitimate	6.54	134	rights	6.21	489	plan	6.43	456
8	corrupt	6.32	212	funding	5.98	312	strategy	6.21	389
9	military	6.21	156	protest	5.87	267	economic	5.98	534
10	pressure	5.98	178	crisis	5.76	312	stability	5.76	312

As shown in Table 3, the collocate data reveal three highly differentiated associative networks surrounding the lemma 'Government'. The PAK sub-corpus profile is dominated by negative collocates that construct government as a fundamentally contested and morally compromised institution: crackdown (MI = 8.76), unconstitutional (MI = 8.43), failed (MI = 7.89), detained (MI = 7.12), and corrupt (MI = 6.32). The concentration of negative modifiers and process-nominals in the 'Government' collocate profile far more distinct than in any international sub-corpus which indicates the sustained critical engagement of Pakistani media with government legitimacy during the 2023 political crisis. This stance does not simply reflect the LW sub-corpus' oppositional collocational profile (organised around austerity, and inequality): while the LW profile constructs government as failing in its distributive responsibilities, the PAK profile constructs government as failing in its constitutional and moral responsibilities, that is, a qualitatively different ideological

criticism that goes beyond economic policy to challenge the legitimacy of the governing authority itself. This distinction maps directly onto Van Dijk's (1998) account of the ideological square: the PAK sub-corpus deploys the negative other-representation strategy not against an out-group constituency (as the RW does with migrants, or the LW does with corporations) but against state power itself by overturning the conventional alignment between mainstream media and governmental authority.

The semantic prosody analysis, summarised in Table 4, extends these findings to a set of politically salient lemmas across the PAK and international sub-corpora, and provides a more clear account of how prosodic meaning (the affective orientation that lexical items acquire through their habitual collocational environments) operates in Pakistani versus international headline discourse.

Table 4: Semantic Prosody Analysis of Selected Lemmas across Sub-Corpora

<b>Lemma</b>	<b>Corpus</b>	<b>Positive Co-text (%)</b>	<b>Negative Co-text (%)</b>	<b>Dominant Prosody</b>	<b>Primary Domain</b>	<b>Ideological Function</b>	<b>Frame Type</b>
<b>accountability</b>	PAK	34.20	65.80	Negative	Corruption allegations; judicial process	Opposition politicians; institutions	Critical exposure
<b>stability</b>	PAK	67.30	32.70	Positive	Economic policy; security narrative	State authority claims	Legitimation of incumbency
<b>stability</b>	RW	71.20	28.80	Positive	Conservative policy outcomes	Government economic record	Self-validation frame
<b>Crisis</b>	PAK	18.40	81.60	Negative	Political turmoil; economic collapse	PTI events; IMF negotiations	Alarm/urgency frame
<b>Crisis</b>	LW	31.20	68.80	Negative	Systemic inequality; housing shortage	Structural causes	Structural critique frame
<b>crisis</b>	RW	14.30	85.70	Negative	Migration; border security	Threat/numbers scale	Threat-security frame
<b>reform</b>	PAK	45.60	54.40	Negative	Economic restructuring (IMF conditions)	IMF; government	External imposition frame
<b>reform</b>	LW	78.40	21.60	Positive	Social policy; public services	Progressive agenda	Aspirational change frame
<b>accountability</b>	LW	82.30	17.70	Positive	Corporate governance; elections	Democratic norms	Democratic accountability
<b>military</b>	PAK	56.70	43.30	Mixed	Security operations; political intervention	State institution	Power/legitimacy duality

As Table 4 shows, the semantic prosody data demonstrate several key findings. The lemma *accountability* presents the most striking cross-sub-corpus difference. In the LW sub-corpus, *accountability* carries predominantly positive prosody (82.3% positive co-text) which functions within a frame of democratic governance in which holding power to account is a civic virtue. In the PAK sub-corpus, by contrast, *accountability* carries predominantly negative prosody (65.8% negative co-text), a pattern that initially appears counterintuitive but is explained by concordance analysis: in the PAK corpus, *accountability* most frequently appears in co-texts describing the National Accountability Bureau, a prosecutorial body that has been widely criticised in Pakistani political discourse as an instrument of political persecution rather than genuine anti-corruption enforcement. The identical lexeme thus carries almost inverse prosodic orientations in the PAK and LW corpora, because the institutional referent, and the power relations it embeds, differs fundamentally between Pakistani and British political contexts. This finding extends Stubbs's (2001) account of semantic prosody by indicating that prosodic polarity is not merely a property of individual lexical items but is constituted through the interaction of lexis with specific national political vocabularies and institutional referential contexts.

The prosody of *reform* across sub-corpora reveals a further dimension of Pakistani-specific framing. In the LW corpus, *reform* carries strongly positive prosody (78.4%) which functions as an aspirational marker of progressive social change. In the PAK corpus, however, *reform* is negatively prosodic (54.4% negative co-text), primarily because it collocates with IMF-mandated economic restructuring (e.g., austerity measures, subsidy removal, privatization) experienced by Pakistani citizens as externally imposed suffering rather than domestically driven improvement. The identical word thus encodes ambition in the LW corpus and external compulsion in the PAK corpus which is a finding that has significant implications for cross-national media comparison: studies that assume semantic equivalence across national corpora will systematically misread the ideological function of politically contested lexemes when comparing Global North and Global South media systems. This conclusion is consistent with Akbar and Saba's (2026) findings that Pakistani and British press discourse exhibit systematic linguistic variation across comparable news domains.

The prosody of *military* in the PAK corpus is uniquely complex: it achieves mixed prosody (56.7% positive, 43.3% negative co-text), and reflects the uncertain positioning of military power in Pakistani public discourse. Concordance analysis reveals two distinct prosodic domains: positive co-text clustering around security operations ('military operation', 'forces successfully') and negative co-text clustering around political intervention ('military pressure', 'establishment interference'). This dual prosody is consistent with Sabahat et al.'s (2025) finding that Pakistani media discourse on political crises reflects a fundamental tension between the legitimacy of military security functions and the contested legitimacy of military political influence, which is a tension that is structurally specific to Pakistan's civil-military relations and has no direct parallel in the LW or RW international sub-corpora.

### **3.3 Transitivity, Agency, and Evaluation**

Table 5 presents the transitivity and agency attribution analysis across all four sub-corpora based on the Python-based annotation of active/passive voice constructions and agent/patient grammatical roles ( $\kappa=0.82$ ).

Table 5: Transitivity and Agency Attribution Patterns across Sub-Corpora

Sub-Corpus	Active Voice (%)	Passive Voice (%)	Agent Foregrounding Rate	Agent Suppression Rate (%)	Primary Context of Agent Suppression (%)	Typical Suppressed Agent(s)
<b>Pakistani (PAK)</b>	72.30	27.70	41.2% (Political leaders)	78.60	67.40	Military, courts, state institutions
<b>Left-Wing (LW)</b>	54.30	45.70	62.4% (Corporate actors)	31.20	47.80	Corporations, governments
<b>Right-Wing (RW)</b>	71.20	28.80	21.3% (Corporate actors)	67.40	22.10	Protesters, migrants, unions
<b>International (INT)</b>	68.90	31.10	34.6% (Institutional actors)	51.30	38.90	Unnamed officials, agencies

As Table 5 shows, the transitivity data reveal that the PAK sub-corpus exhibits the highest rate of active voice constructions (72.3%) of all four sub-corpora, a figure comparable to the RW sub-corpus (71.2%) but achieved through fundamentally different agent patterns. In the RW sub-corpus, high active voice rates reflect the foregrounding of corporate and governmental actors as rational economic decision-makers. In the PAK sub-corpus, high active voice rates reflect the foregrounding of political and military actors as agents of politically opposed actions (e.g., arrests, crackdowns, constitutional rulings) in which agent clarity is itself an ideological choice. The framing of PTI-related arrests exemplifies this pattern: where a Pakistani headline reads ‘Police Detain 200 PTI Workers as Crackdown Widens’ (foregrounding police agency and implying systematic state action), the grammatically equivalent passive ‘200 Detained in PTI Crackdown’ would suppress police agency and reduce the apparent intentionality of the suppression. The PAK sub-corpus preference for active constructions in political arrest contexts thus encodes a critical stance toward state action that is achieved through grammatical foregrounding rather than explicit evaluative language which is consistent with Fowler’s (1991) account of transitivity as a covert ideological resource.

The agent suppression analysis reveals a further PAK-specific pattern. At 78.6% agent suppression rate for opposition and civil society actors, the PAK sub-corpus shows a strong tendency to present political opposition members, civil society voices, and ordinary citizens as patients of state action rather than as autonomous social agents. This passivisation pattern (in which the same sub-corpus that foregrounds state agency in active constructions simultaneously suppresses civil society agency through passive constructions) creates a systematic discursive asymmetry in which governmental and military actors appear as the primary movers of political events while opposition and civic actors appear as their passive recipients. This pattern instantiates Van Dijk’s (1998) ideological square in a particularly pointed form: the PAK sub-corpus distribution of active and passive constructions encodes a power asymmetry between state and civil society that reflects, and potentially reinforces, the actual power asymmetries of Pakistani political life during 2023.

Comparison with the international sub-corpora contextualises this finding significantly. The LW sub-corpus shows a reversed passivisation pattern: corporate actors are suppressed at a 62.4% rate in economic reporting (producing constructions like ‘500 Jobs Lost as Factory Closes’ that obscure corporate responsibility), while social actors are grammatically foregrounded in active

constructions. This reflects the LW corpus systematic alignment with labour and civil society perspectives at the expense of corporate accountability which is a finding consistent with Baker et al.'s (2008) account of how corpus analysis can reveal ideological patterns that are invisible to casual reading. The RW sub-corpus shows a further variation: high active voice rates for government and economic actors (representing them as purposeful agents) combined with higher passivisation of migrant and protest actors. The PAK pattern thus differs from both international patterns not merely in its specific agency attributions but in the structural relationship between active/passive distribution and political alignment where the LW passivises those with economic power and the RW passivises those who challenge political or regional boundaries, the PAK sub-corpus passivises those who challenge government authority which reveals a distinct power-linguistic structure specific to the Pakistani political context. These findings build on and extend Khalifah et al.'s (2026) comparable transitivity analysis of Al Jazeera's bilingual conflict reporting in which they report systematic difference between Arabic and English platforms in the passivisation of Palestinian casualty descriptions, and demonstrates that transitivity choices are among the most ideologically loaded and audience-related features of news discourse.

Almohawes's (2026) corpus-based study of Saudi Vision 2030 in international English-language media provides a productive point of comparison for the PAK sub-corpus economic discourse. Where Almohawes (2026) finds that international coverage of Saudi Vision 2030 constructs the Saudi state as an active, agentive reformer through verbs like 'targeted', 'introduced', and 'diversify' (collocates: 'goals' MI 5.25, 'objectives' MI 6.50, 'diversification' MI 6.51), the PAK sub-corpus constructs the Pakistani state primarily as the patient of external economic disciplines ('IMF demands', 'rupee falls', 'inflation reaches') with the rare active-voice constructions available to Pakistani governmental actors clustering around political suppression rather than economic agency. This comparison emphasizes the extent to which the transitivity patterns of news discourse reflect not merely editorial ideology but the basic political and economic realities being reported upon which is a findings that, as Sabahat et al. (2025) report, significantly shapes public perception and trust in political institutions.

#### **4. Conclusion**

This study has presented the first CADS study of Pakistani English-language news headline discourse within an international comparative framework by taking a corpus of 20,000 headlines (271,755 tokens) from twelve outlets across four ideologically differentiated sub-corpora. Through a multi-method analytical protocol (i.e., log-likelihood keyness analysis, MI-based collocate extraction using AntConc 4.2 and Sketch Engine, semantic prosody analysis, and semi-automated transitivity annotation) this study has produced three categories of finding corresponding to its three research questions.

In relation to Research Question 1, keyword analysis presented in Table 2 reveals that Pakistani news headline language is organised around a lexical domain that has no direct equivalent in international sub-corpora: a cluster of political-institutional crisis vocabulary (accountability, crackdown, military, detained, establishment) which indicates the specific conditions of Pakistani civil-military relations, judicial contestation, and political repression in 2023. This vocabulary is accompanied by a distinctive economic crisis lexicon (IMF, rupee, inflation) that constructs the Pakistani state as the patient of external economic discipline rather than the autonomous agent of

economic choice, a fundamentally different framing of economic responsibility from either the LW (distributive failure) or RW (anti-growth constraint) patterns. In relation to Research Question 2, collocates and semantic prosody analyses demonstrated in Tables 3 and 4 that the identical lexemes carry systematically different prosodic orientations across national contexts: accountability is positively prosodic in the LW corpus (civic virtue) but negatively prosodic in the PAK corpus (instrument of political persecution); reform encodes aspirational change in the LW corpus but externally imposed austerity in the PAK corpus; military achieves a uniquely ambivalent dual prosody in Pakistan which indicates the opposed legitimacy of military power in Pakistani political culture. These findings demonstrate that semantic prosody is constituted not only through lexical co-occurrence patterns but through the interaction of lexis with specific national political vocabularies, institutional referents, and power relations. In relation to Research Question 3, transitivity analysis presented in Table 5 reveals that the PAK sub-corpus high active voice rate encodes a critical stance toward state action through grammatical foregrounding of government and military agency, while simultaneously suppressing civil society and opposition agency through passivisation, that is, a power-linguistic structure that is distinct from both LW and RW international patterns and reflects the specific asymmetries of Pakistani civil-military relations in 2023.

#### **4.1 Research Implications**

The findings carry three key implications. For media discourse research, the study demonstrates that semantic equivalence cannot be assumed across national corpora as lexemes like accountability and reform carry opposite prosodic orientations in Pakistani versus international contexts. Comparative corpus studies must therefore focus on nationally specific political vocabularies, not merely statistical co-occurrence patterns. For journalism practice and media literacy, this study provides empirically grounded tools for teaching headline reading by specifying the lexical and grammatical mechanisms through which ideological framing operates at the headline level. For political communication, the observation of systematic passivisation of civil society actors in Pakistani headlines reveals that ideological work extends beyond explicit evaluative commentary to the grammatical structuring of agency itself with direct implications for understanding how media language may reinforce political power asymmetries.

#### **4.2 Directions for Future Research**

Four future directions follow in this study. First, an Urdu-language corpus extension would enable analysis of whether ideological patterns replicate across Pakistan's majority-language press and enable bilingual comparison. Second, diachronic analysis across multiple years would distinguish robust framing characteristics from episodic crisis-specific patterns. Third, multimodal extension incorporating image analysis would provide a more complete account of framing resources in Pakistani digital journalism. Fourth, audience reception studies combining textual analysis with reader data would reveal how different political affiliations process the specific framing devices identified here by completing the cycle from production to interpretation.

**Funding:** This study was not funded in any shape or form by any party.

**Conflict of Interest:** The author declares that he has no conflict of interest.

**Bio-note:**

**Ali Raza** is an M.Phil. Scholar in the Department of English at the University of Okara, Pakistan. His research interests include corpus linguistics, critical discourse analysis, media discourse, political communication, corpus-assisted discourse studies, and computational approaches to language analysis. His current research focuses on the ideological representation of political events in digital news media through corpus-assisted critical discourse analysis.

**Dr. Muhammad Ahmad** is a Visiting Lecturer in the Department of English at the University of Okara, Pakistan. His research specializes in corpus linguistics, applied linguistics, discourse analysis, media discourse, academic writing, and English language studies. He has published extensively in national and international peer-reviewed journals, particularly in the areas of corpus-based linguistic analysis, multidimensional analysis, and media discourse.

**Kainat Zulfiqar** is an M.Phil. Scholar in the Department of English at the University of Okara, Pakistan. Her research interests include corpus linguistics, critical discourse analysis, sociolinguistics, media studies, gender discourse, and digital communication. Her academic work focuses on the application of corpus-based methods to the investigation of language, ideology, and media representation.

**References**

- Ahmad, M., Mahmood, M. A., Mahmood, M. I., & Siddique, A. R. (2019). Use of modal verbs as stance markers in Pakistani English newspaper editorials. *Online Journal of Communication and Media Technologies*, 9(1), 1–14. <https://doi.org/10.29333/ojcm/5722>
- Ahmad, M., Mahmood, M. A., & Farukh, A. (2020). Use of modals as stance markers: A corpus-based study on Pakistani English newspaper editorials. *Asia Pacific Media Educator*, 30(1), 108–125. <https://doi.org/10.1177/1326365X20945424>
- Akbar, M., & Saba, N. (2026). Linguistic variation in Pakistani and British crime press reportage: A multidimensional analysis. *Linguistic Forum*, 8(1), 1–20. <https://linguisticforum.com/index.php/ling/article/view/311>
- Akhter, N., & Shah, S. K. (2024). An analysis of how women's rights are framed in Pakistani English newspapers: A corpus-based discourse study. *Linguistic Forum*, 6(4), 60–78. <https://www.linguisticforum.com/index.php/ling/article/view/276>
- Alluhaidah, M. (2023). A critical discourse analysis of Al Jazeera's online coverage of the war in Yemen before and after the 5 June 2017 Gulf crisis. *Discourse & Communication*, 17(5), 553–589. <https://doi.org/10.1177/17504813231163769>
- Almohawes, M. (2026). A corpus-based critical discourse analysis of news on the web representing Saudi Vision 2030. *Frontiers in Communication*, 11, 1–16. <https://doi.org/10.3389/fcomm.2026.1801652>

- Anthony, L. (2022). *AntConc* (Version 4.2.4) [Computer software]. Waseda University. <https://laurenceanthony.net/software.html>
- Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.
- 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>
- Baker, P., Gabrielatos, C., & McEnery, T. (2013). *Discourse analysis and media attitudes: The representation of Islam in the British press*. Cambridge University Press.
- Bednarek, M. (2016). Investigating evaluation and news values in news items that are shared via social media. *Corpora*, 11(2), 237–257. <https://doi.org/10.3366/cor.2016.0093>
- Bednarek, M., & Caple, H. (2017). *The discourse of news values: How news organizations create newsworthiness*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190653934.001.0001>
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.
- Brookes, G., & McEnery, T. (2019). The utility of topic modelling for discourse studies: A critical evaluation. *Discourse Studies*, 21(1), 3–21. <https://doi.org/10.1177/1461445618814032>
- Caldas-Coulthard, C. R., & Coulthard, M. (Eds.). (1996). *Texts and practices: Readings in critical discourse analysis*. Routledge.
- Church, K., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1), 22–29.
- Conboy, M. (2007). *The language of the news*. Routledge.
- Dor, D. (2003). On newspaper headlines as relevance optimizers. *Journal of Pragmatics*, 35(5), 695–721. [https://doi.org/10.1016/S0378-2166\(02\)00134-0](https://doi.org/10.1016/S0378-2166(02)00134-0)
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51–58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- Entman, R. M. (2007). Framing bias: Media in the distribution of power. *Journal of Communication*, 57(1), 163–173. <https://doi.org/10.1111/j.1460-2466.2006.00336.x>

- Fairclough, N. (1995). *Media discourse*. Edward Arnold.
- Fatima, M., Siddique, A. R., Ahmad, M., & Mahmood, M. A. (2023). Exploring linguistic variation in Pakistani English newspaper editorials through multidimensional analysis. *Newspaper Research Journal*, 44(4), 425–451. <https://doi.org/10.1177/07395329231194005>
- Fowler, R. (1991). *Language in the news: Discourse and ideology in the press*. Routledge.
- Gabrielatos, C. (2018). Keyness analysis: Nature, metrics and techniques. In C. Taylor & A. Marchi (Eds.), *Corpus approaches to discourse: A critical review* (pp. 225–258). Routledge.
- Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Harvard University Press.
- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). Edward Arnold.
- Herman, E. S., & Chomsky, N. (2021). Manufacturing consent. In L. Chorbajian (Ed.), *Power and inequality: Critical readings for a new era* (pp. 198–206). Routledge. <https://doi.org/10.4324/9781315201511>
- Hunston, S., & Thompson, G. (Eds.). (2000). *Evaluation in text: Authorial stance and the construction of discourse*. Oxford University Press.
- Khalifah, L., Altakhaineh, A. R. M., & Za'rour, R. (2026). A critical discourse analysis of ideological positioning in Al Jazeera's bilingual coverage of the 2023–2024 Gaza–Israel war. *Cogent Arts & Humanities*, 13(1), 2–17. <https://doi.org/10.1080/23311983.2026.2664226>
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1(1), 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- Mullainathan, S., & Shleifer, A. (2005). The market for news. *American Economic Review*, 95(4), 1031–1053. <https://doi.org/10.1257/0002828054825619>
- 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>
- Richardson, J. E. (2007). *Analysing newspapers: An approach from critical discourse analysis*. Palgrave Macmillan.

- Sabahat, S., Iqbal, N., Gulzar, M. F., & Arshad, A. A. (2025). Framing and ideology: A critical discourse analysis of news media representation in political crises in Pakistan. *The Asian Bulletin of Big Data Management*, 5(3), 46–62. <https://doi.org/10.62019/1y8f7j41>
- Siddique, A. R., Mahmood, M. A., & Ahmad, M. (2022). Diachronic variation in the language of Pakistani English newspapers: A multidimensional analysis. *Corporum: Journal of Corpus Linguistics*, 5(2), 109–127. <https://journals.au.edu.pk/ojsrcr/index.php/crc/article/view/239>
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Blackwell.
- Van Dijk, T. A. (1988). *News analysis: Case studies of international and national news in the press*. Lawrence Erlbaum.
- Van Dijk, T. A. (1998). *Ideology: A multidisciplinary approach*. Sage.
- Van Dijk, T. A. (2009). News, discourse, and ideology. In K. Wahl-Jorgensen & T. Hanitzsch (Eds.), *The handbook of journalism studies* (pp. 191–204). Routledge.