A Meta-Analysis of Corpus-Based Approaches to CDA: Pakistani and International Perspectives (2015–2025)

ABSTRACT:

This article presents a comprehensive meta-analysis of corpus-based approaches to Critical Discourse Analysis (CDA) by synthesizing empirical research from Pakistani and international contexts between 2015 and 2025. Critical Discourse Analysis is a multidisciplinary research paradigm that explores how language constructs and reflects power, ideology, and inequality. When combined with corpus linguistics (CL), CDA gains methodological rigor by integrating computational tools, such as concordance, collocation, frequency analysis, and keyword analysis. The present study systematically reviews ten research articles, five from Pakistan and five from global contexts, to examine methodological designs, corpus compilation strategies, analytical frameworks, and thematic foci. Findings reveal that Pakistani studies often emphasize political, educational, and media discourses but face limitations due to underdeveloped corpora and a lack of digitization in regional languages. International research, by contrast, benefits from advanced corpus resources and large datasets but risks cultural misinterpretation and excessive reliance on quantitative techniques. The study contributes to theoretical advancement by mapping the synergy between CDA and corpus linguistics, methodological innovation through identifying strengths and gaps, and pedagogical utility by offering a synthesized framework for future research. Implications include the need for multilingual corpus development in Pakistan, qualitative–quantitative integration in international studies, and cross-cultural comparisons to enhance validity and inclusivity. Ultimately, corpus-based CDA demonstrates considerable potential for analyzing ideology, representation, and power across a wide range of discourses, though its effectiveness depends on balancing computational rigor with critical interpretation.

AUTHORS:

Rubeena Ghaffar
Lecturer, Lahore Garrison University

Kalsoom Afzaal
Lecturer, Minhaj University Lahore

KEYWORDS:
Corpus Linguistics, Critical Discourse Analysis, Ideology, Meta-Analysis, Discourse Studies

Introduction

Critical Discourse Analysis (CDA) has emerged as a significant approach in applied linguistics and discourse studies for examining the relationship between language, power, and ideology. Rooted in critical theory, CDA challenges the notion of language as a neutral medium of communication, instead treating it as a social practice that both reflects and shapes societal structures (Fairclough, 1995; van Dijk, 2009; Wodak, 2001). By interrogating how texts construct and perpetuate ideologies, CDA highlights the often-unseen mechanisms of dominance, exclusion, and inequality embedded within linguistic choices. Scholars such as Norman Fairclough, Teun A. van Dijk, and Ruth Wodak have developed influential frameworks for CDA that continue to guide contemporary research across political, media, educational, and cultural contexts.

In parallel, corpus linguistics (CL) has developed as a methodological paradigm that employs computational tools to analyze large bodies of authentic texts systematically (Baker, 2006; Stubbs, 1997). CL techniques, including frequency counts, collocation analysis, concordancing, and keyword extraction, allow researchers to uncover recurring patterns of language use across extensive datasets. Such patterns can reveal ideological tendencies or social positioning that may not be evident in isolated texts. While CDA provides a critical interpretative framework, CL offers methodological rigor and replicability. When combined, corpus-based CDA thus represents a powerful synergy that balances quantitative analysis with qualitative interpretation (Baker et al., 2008; Mautner, 2009).

The convergence of CDA and CL has given rise to a rapidly expanding body of research, particularly in contexts where large-scale textual data is available. This methodological integration has been employed in studies of political discourse, media representations, educational policy, health communication, and online interactions. For example, corpus-assisted CDA has been applied to investigate media portrayals of refugees, ideological constructions in political speeches, and identity representations in educational curricula (Baker, 2010; Vessey, 2017). By foregrounding both patterns and their social implications, this combined approach ensures that discourse studies remain both empirically grounded and critically engaged.

Despite these advances, gaps remain in how corpus-based CDA is conceptualized and applied across different socio-political contexts. In Pakistan, CDA has gained prominence as a framework for interrogating issues of power, ideology, and representation in education, politics, and media (Mushtaq & Zahid, 2016). However, Pakistani scholarship often faces limitations such as the scarcity of digitized corpora, underrepresentation of regional languages, and restricted access to large-scale datasets. Consequently, much of the work relies on smaller samples or mixed-methods approaches that integrate surveys and

interviews with textual analysis. By contrast, international scholarship benefits from extensive digital corpora, advanced analytical tools, and established traditions in corpus linguistics. Yet, international studies sometimes risk overreliance on quantitative measures, such as frequency counts, at the expense of deeper contextual and cultural interpretations (Baker, 2012; Mautner, 2009).

A meta-analysis of corpus-based CDA is therefore both timely and necessary. Meta-analysis, as a systematic review method, synthesizes empirical studies to identify patterns, methodological trends, and research gaps (Cooper, 2010). In the context of CDA and CL, such an approach allows for an evaluation of how corpus-based methods are operationalized across diverse settings, the extent to which they contribute to uncovering ideology and power, and what limitations they face. Moreover, examining both Pakistani and international contributions enables a comparative perspective that highlights how local and global research trajectories inform each other. This comparative synthesis not only deepens our understanding of methodological strengths and weaknesses but also provides a roadmap for future scholarship.

The present study aims to review ten selected works, five from Pakistan and five from international contexts, published between 2015 and 2025. These studies span diverse themes, including Islamophobia in media discourse, ideological shifts in Pakistani political rhetoric, language ideologies in educational policy, editorial constructions of terrorism, global representations of China amid COVID-19, online hate speech in Europe, and ideological framing of war discourses. By analyzing methodological choices, corpus compilation strategies, analytical tools, and theoretical frameworks, this article seeks to identify both convergences and divergences in corpus-based CDA practices.

The significance of this study is threefold. Theoretically, it advances discourse studies by mapping the operationalization of corpus linguistics within CDA frameworks, showing how computational tools and interpretive models complement one another. Methodologically, it provides insights into the strengths and gaps of current approaches, guiding future researchers toward more robust and replicable designs. Practically, it highlights the pedagogical and social utility of corpus-based CDA by showing its relevance to issues of politics, media, education, and public discourse. In contexts such as Pakistan, where multilingualism and under-digitization pose significant challenges, this synthesis highlights the urgent need to develop open-access corpora and expand the scope of research beyond English-dominant texts.

In sum, this article situates itself at the intersection of CDA and corpus linguistics, offering a meta-analytic synthesis that compares Pakistani and international scholarship. By doing so, it aims to contribute to both the theoretical advancement and practical relevance of discourse studies. The subsequent sections present a review of key theoretical frameworks (Section 3), outline the methodology of the meta-analysis

(Section 4), provide an in-depth analysis and discussion of selected studies (Section 5), and conclude with findings, implications, and recommendations for future research (Sections 6–7).

Literature Review

Critical Discourse Analysis (CDA) emerged in the late 20th century as a response to traditional linguistics and discourse studies that often treated language as an autonomous system, detached from its socio-political functions. Scholars such as Norman Fairclough, Teun A. van Dijk, and Ruth Wodak pioneered CDA to foreground the role of discourse in constructing and sustaining social inequalities (Fairclough, 1992, 1995; van Dijk, 1997, 2009; Wodak, 2001). Unlike descriptive approaches to discourse, CDA is explicitly normative and interventionist: it not only analyzes language but also seeks to reveal and challenge ideological manipulation, dominance, and exclusion.

CDA's central assumption is that discourse is both socially shaped and socially constitutive. Language choices are never neutral but reflect ideological positions that reproduce or resist power. For instance, the framing of immigrants as "illegal" in political discourse not only describes a legal status but also constructs a delegitimizing identity that justifies exclusion (van Dijk, 1993). Similarly, the use of passive constructions in policy documents can obscure agency and accountability, thereby legitimizing dominant structures (Fairclough, 2001). Thus, CDA treats texts as sites of struggle where competing ideologies are discursively constructed.

Although CDA is unified by its critical orientation, it encompasses several distinct theoretical frameworks. Three of the most influential are Fairclough's dialectical-relational approach, van Dijk's socio-cognitive model, and Wodak's discourse-historical approach.

Fairclough (1992, 1995) conceptualizes discourse as a form of social practice embedded within a dialectical relationship with social structures. His three-dimensional model of CDA includes (a) textual analysis, focusing on vocabulary, grammar, and cohesion; (b) discursive practice, examining how texts are produced, distributed, and consumed; and (c) social practice, situating discourse within wider ideological and institutional contexts. This model has been widely adopted for analyzing media texts, political speeches, and educational policies, as it allows for multi-level analysis that connects micro-linguistic details with macro-social structures.

Van Dijk (1997, 2009) emphasizes the role of cognition in mediating between discourse and society. He argues that mental representations, or "social cognition," link individual text processing with collective ideologies. Central to his model is the concept of the "ideological square," which structures discourse

around in-group favoritism and out-group derogation: emphasizing "our" good qualities, deemphasizing "our" bad qualities, emphasizing "their" bad qualities, and deemphasizing "their" good qualities. This model has been particularly influential in political discourse analysis, where strategies of positive self-presentation and negative other-presentation are frequently observed.

Wodak (2001, 2015) situates discourse within its historical and intertextual contexts, emphasizing the importance of temporality, interdiscursivity, and socio-political events. The DHA integrates linguistic analysis with historical knowledge to reveal how discourses evolve across time and how they draw on prior texts. This approach has been extensively applied to studies of nationalism, racism, and European Union policies, where understanding intertextuality and historical continuity is crucial. Together, these frameworks demonstrate the theoretical richness of CDA. While Fairclough emphasizes dialectics, van Dijk highlights cognition, and Wodak stresses history, they all converge on the view that discourse is ideological and constitutive of social power.

Despite its impact, CDA has faced critiques. Stubbs (1997) argued that CDA sometimes relies on "cherry-picked" examples rather than systematic datasets, raising concerns about subjectivity and replicability. Similarly, Schegloff (1997) criticized CDA for insufficient attention to interactional detail compared to conversation analysis. Moreover, CDA's commitment to critique has been questioned for blurring the line between analysis and advocacy (Widdowson, 1998). These critiques highlight the need for methodological transparency and systematicity, which corpus linguistics can provide.

Corpus linguistics (CL) is a methodological approach that employs computational tools to analyze large, principled collections of authentic language data (Sinclair, 1991; McEnery & Hardie, 2012). Unlike impressionistic analyses, CL enables researchers to identify recurring linguistic patterns across extensive datasets, thereby enhancing reliability and representativeness.

Key techniques in CL include:

Frequency analysis: Which identifies the most common words or structures in a corpus, revealing thematic emphases.

Collocation analysis: Which examines the company words kept, highlighting semantic associations and evaluative tendencies (Stubbs, 2001).

Concordancing: Which allows for the examination of words in context, illuminating discourse prosody (Louw, 1993).

Keyword analysis: Which compares corpora to identify statistically significant words that mark

discursive salience (Scott, 1997).

These tools provide an empirical foundation for discourse studies by enabling the detection of subtle patterns and ideological framing. For example, Baker (2006) demonstrated how keywords and collocations can reveal latent biases in media discourse about refugees.

The integration of CDA and CL, often termed "corpus-assisted discourse analysis" (Baker et al., 2008) or "corpus-based CDA" (Mautner, 2009), represents a methodological synergy. While CDA provides a critical framework for interpreting power and ideology, CL offers systematic tools for detecting linguistic patterns across large datasets. Together, they combine the strengths of qualitative interpretation and quantitative rigor.

Baker et al. (2008), in their seminal study of UK press discourses on refugees, combined keyword analysis with CDA to show how refugees were consistently associated with burden metaphors and criminality. Similarly, Gabrielatos and Baker (2008) analyzed British press coverage of Islam, finding patterns of negative framing through collocation networks. These studies illustrate how corpus-based CDA allows researchers to uncover systematic biases that might otherwise remain hidden.

The benefits of this integration include:

- **Replicability:** Corpus-based methods provide transparent procedures that can be independently verified.

- **Representativeness:** Large corpora capture broader trends beyond isolated texts.

- **Sensitivity to patterning:** CL techniques reveal recurrent lexical and grammatical patterns that reflect ideological positioning.

However, challenges remain. Overreliance on quantitative metrics may obscure contextual nuance, leading to superficial interpretations (Mautner, 2009). Additionally, corpus compilation itself is not neutral; decisions about which texts to include shape the results (McEnery & Hardie, 2012). Thus, corpus-based CDA must balance computational rigor with critical reflexivity.

In Pakistan, CDA has been widely applied to analyze political speeches, educational curricula, and media representations (Mushtaq & Zahid, 2016; Abbas & Shehzad, 2018). Yet, most studies rely on small-scale datasets due to the lack of digitized corpora and computational infrastructure. Consequently, corpus-based CDA remains underdeveloped, with only a few studies attempting systematic corpus construction or large-scale computational analysis. This limitation constrains the representativeness and replicability of

findings. Nonetheless, Pakistani research contributes valuable insights into local ideological struggles, particularly around issues of Islamophobia, language policy, and political discourse.

This article adopts a "meta-analytic framework" that synthesizes corpus-based CDA studies from both Pakistani and international contexts. Theoretically, it aligns with Fairclough's dialectical-relational view of discourse as a form of social practice, while also drawing on van Dijk's socio-cognitive insights into ideological framing and Wodak's discourse-historical emphasis on context. Methodologically, it recognizes the value of CL tools for ensuring systematicity while also acknowledging their limitations.

By situating itself at the intersection of CDA and CL, this study underscores the need for methodological pluralism. It argues that corpus-based CDA is most effective when it integrates quantitative detection of patterns with qualitative interpretation of their ideological significance. Moreover, in multilingual and under-digitized contexts such as Pakistan, the development of open-access corpora and context-sensitive frameworks is crucial for advancing discourse studies.

Research Methodology

The present study employs a "meta-analysis design" to synthesize research on corpus-based approaches to Critical Discourse Analysis (CDA). Meta-analysis, broadly defined, refers to the systematic collection, evaluation, and synthesis of empirical studies on a specific research topic (Cooper, 2010). In discourse studies, meta-analysis enables researchers to move beyond individual case studies to identify broader patterns, methodological trends, and conceptual frameworks. Given the increasing but fragmented nature of research combining corpus linguistics (CL) with CDA, this design is particularly appropriate for mapping the field across local (Pakistani) and global (international) contexts.

The analysis is delimited to the period "2015–2025" to capture recent developments in corpus-based CDA and to ensure that both Pakistani and international contributions reflect contemporary methodological and theoretical advancements. A total of "10 studies" were selected: five from Pakistani journals and repositories, and five from internationally recognized outlets. The inclusion of both contexts allows for comparative insights into how corpus-based CDA has been conceptualized and applied in different socio-political and cultural environments.

The search strategy was multi-layered to ensure comprehensive coverage. For Pakistani studies, databases such as the "Pakistan Research Repository (PRR)" and the Higher Education Commission's "HEC Journal Recognition System (HJRS)" were used. Searches were also conducted in university-based journals such as the 'Journal of Social Sciences and Humanities' (University of Karachi) and 'Pakistan Journal of Social Sciences' (Bahauddin Zakariya University). For international studies, databases such as

"Elsevier", "SAGE Journals", "Taylor & Francis Online", "Cambridge University Press", and "Springer" were consulted. In addition," Google Scholar" and publisher websites were searched to capture grey literature and open-access publications.

The following keywords were employed in various combinations: "corpus linguistics," "critical discourse analysis," "CDA," ideology," "political discourse," and "media discourse." Boolean operators (AND/OR) were applied to refine the searches.

To ensure consistency, strict inclusion and exclusion criteria were applied.

Inclusion criteria:

- The study explicitly combined corpus linguistics methods with CDA frameworks.
- The publication was peer-reviewed or published by a recognized academic outlet.
- The study was empirical, reporting original data rather than purely theoretical reflections.
- The focus was on discourse in political, media, educational, or social contexts.
- The publication date was between 2015 and 2025.

Exclusion criteria:

- Studies that used corpus linguistics without a CDA framework.
- Purely theoretical or conceptual articles without empirical data.
- Studies focusing exclusively on technical corpus methods with no critical analysis of ideology or power.
- Publications not available in English.

Applying these criteria resulted in a final sample of **10 studies** that met all requirements.

For each selected study, relevant information was extracted using a structured data sheet. The following categories were recorded:

- Bibliographic details (author, year, title, source)
- Corpus type and size (e.g., newspaper articles, speeches, policy documents, online texts)
- Analytical tools (e.g., AntConc, Sketch Engine, SPSS for quantitative triangulation)
- CDA framework employed (Fairclough's three-dimensional model, van Dijk's socio-cognitive model, Wodak's DHA, or mixed approaches)
- Methodological design (qualitative, quantitative, or mixed methods)
- Key findings (ideological themes, power relations, discursive strategies)

This systematic extraction ensured comparability across studies while maintaining attention to their specific contexts.

The analysis proceeded in three stages:

- **Descriptive Mapping**

Each study was summarized in terms of corpus characteristics, analytical methods, and CDA framework. This stage provided an overview of the methodological diversity in corpus-based CDA.

- **Comparative Analysis**

Pakistani and international studies were compared to identify convergences and divergences. This comparison considered issues such as corpus availability, linguistic representation, integration of qualitative and quantitative approaches, and thematic foci (e.g., politics, education, media, and health).

- **Synthesis and Interpretation**

Patterns across the ten studies were synthesized to evaluate strengths, limitations, and gaps in the field. For example, while international studies often demonstrated methodological sophistication through large corpora and advanced software, Pakistani studies highlighted the challenges of multilingualism and under-digitization. The interpretation was grounded in the theoretical frameworks of CDA, particularly Fairclough's emphasis on social practice, van Dijk's socio-cognitive model, and Wodak's historical orientation.

Like any meta-analysis, this study faces certain limitations. First, the selection of only ten studies, while manageable for in-depth analysis, may not capture the full global scope of corpus-based CDA. Second, restricting the review to English-language publications potentially excludes important research conducted in regional languages. Third, the reliance on available databases means that unpublished dissertations or internal reports may have been overlooked. Despite these limitations, the chosen sample provides a robust basis for comparative analysis between Pakistani and international research contexts.

As this study is a secondary analysis of published work, no direct ethical approval was required. However, ethical principles were observed by ensuring accurate representation of each study's findings, acknowledging limitations, and providing proper attribution through APA 6th edition referencing.

Data Analysis

The synthesis of the ten selected studies illustrates the diverse ways corpus linguistics has been integrated with Critical Discourse Analysis (CDA) across both Pakistani and international contexts. While

all studies share the aim of uncovering ideological underpinnings in discourse, their methodological choices, corpus types, and theoretical frameworks vary considerably. Pakistani research demonstrates innovation within resource-constrained environments, often compensating for limited corpus availability through mixed-method approaches. In contrast, international studies showcase the advantages of large corpora and advanced computational tools but sometimes risk prioritizing quantitative findings over contextual interpretation.

Pakistani Context Studies

1. Islamophobia in Media Discourse

This study compared how American and Pakistani newspapers discursively construct Islamophobia. Using Fairclough's (1995) sociocultural approach, qualitative micro- and macro-level textual analysis was combined with quantitative data from a survey of 50 respondents analyzed via SPSS. Concordance lines revealed that American newspapers often foregrounded Islam with associations of violence and extremism, while Pakistani newspapers emphasized victimhood and discrimination. The study highlighted CDA's strength in bridging textual analysis with social perceptions, but the small corpus size and reliance on frequency counts limited generalizability. Methodologically, this reflects Pakistan's challenge of corpus digitization but also its willingness to integrate multiple methods to compensate.

2. Diets, Diseases, and Ideological Representation

In a comparative CDA of research articles versus medical association texts on Type 2 diabetes, the researchers found that medical associations tended to obscure responsibility through generalized, ambiguous language. Using Fairclough's three-dimensional model and van Dijk's macro-rules, the analysis revealed that associations discursively shifted responsibility away from structural causes (e.g., food industry practices) toward individual lifestyle choices. This ideological framing aligns with neoliberal health discourses that depoliticize responsibility. By contrast, research articles employed more precise language tied to empirical findings. This study demonstrates how corpus tools such as collocation analysis can reveal hidden ideologies in seemingly neutral scientific discourse.

3. Ideological Shifts in Political Narratives

This study examined the discursive construction of ideology in political speeches by Imran Khan and Nawaz Sharif, focusing on shifts from right-wing to left-wing rhetoric. Using van Dijk's (1997) socio-cognitive model, the analysis identified frequent use of repetition, positive self-representation, and negative other-representation, especially during political crises. The corpus, though limited to speeches available online, provided enough data to reveal ideological shifts over time. The findings underscore how political leaders strategically adapt discourse to changing contexts, illustrating van Dijk's view of ideology as a

flexible cognitive structure reproduced through language.

4. Language Ideologies in Education Policy

A corpus-assisted CDA of 32 Pakistani language policy documents (2000–2020) created a 1.28 million-word corpus, analyzed using frequency, collocation, and concordance techniques. The study revealed a consistent privileging of English and Urdu at the expense of regional languages, reinforcing linguistic hierarchies that marginalize ethnic minorities. Drawing on Fairclough (2013) and Wodak (2001), the analysis showed how policy discourse naturalizes monolingual norms under the guise of modernization. This study is notable for overcoming the challenge of corpus scarcity by compiling a substantial dataset. However, its reliance on official documents risks overlooking grassroots discourses, a limitation acknowledged by the authors.

5. Editorials on Terrorism

This qualitative CDA compared editorials from "Dawn" (Pakistan) and "Hindustan Times"(India) regarding the 2016 surgical strikes. The analysis revealed stark ideological contrasts: Indian discourse emphasized retaliation and national pride, while Pakistani discourse foregrounded peace and de-escalation. Despite the absence of a large corpus, discourse-historical analysis (Wodak, 2015) allowed researchers to situate these texts within broader narratives of national identity and conflict. This study illustrates the flexibility of CDA frameworks in capturing ideological conflict, but also reflects the limitation of a small sample size typical of Pakistani research.

International Context Studies

6. Media Representation of China amid COVID-19

A large-scale corpus (2.89 million words) of "Xinhua News Agency" texts was analyzed to examine China's discursive construction during the pandemic. Using collocation, concordance, and semantic prosody, the study identified China as discursively positioned in three roles: victim, fighter, and cooperative supporter. These representations reinforced ideological narratives of global solidarity and humanitarianism. Wodak's discourse-historical approach provided a framework for linking these discourses to geopolitical strategies. This study exemplifies the strength of corpus-assisted CDA in handling large datasets and identifying systematic framing, though it also risks taking official discourse at face value without incorporating counter-narratives.

7. Hate Speech in German and Polish Social Media

This study investigated hate speech on Facebook by compiling comparable corpora of user-generated comments. Combining corpus linguistics with CDA, it revealed that German users often employed euphemistic or indirect expressions, while Polish users tended toward overtly aggressive language. These

findings confirm van Dijk's (2009) argument that discourse reflects socio-cultural norms. By employing collocation and concordance, the study demonstrated how hate speech is linguistically conditioned by national contexts. Its strength lies in its comparative design, but the reliance on Facebook data may not capture the full spectrum of online hate speech across platforms.

8. CDA of Martin Luther King's Speech

Applying Fairclough's three-dimensional model to King's "I Have a Dream" speech, this study analyzed textual, discursive, and social dimensions. Corpus-assisted tools highlighted recurrent lexical patterns such as "freedom" and "justice," which functioned as cohesive devices to construct an inclusive ideological vision. CDA revealed how stylistic strategies reflected Afro-American struggles for equality. This study underscores the value of corpus tools in systematically analyzing iconic texts but is limited by its focus on a single speech rather than a broader corpus. Nonetheless, it demonstrates CDA's relevance in uncovering the socio-political implications of stylistic choices.

9. CDA of Song Lyrics ("Fanaa Ho" from Frozen II)

Using Fairclough's model, this study applied CDA to song lyrics, demonstrating the adaptability of the framework to non-traditional texts. Collocational and concordance analyses revealed implicit ideologies of power, gender, and hierarchy embedded in the lyrics. The study argued that popular culture texts, often overlooked in CDA, play a significant role in shaping social imaginaries. While innovative, the study's small dataset limits generalizability. Nevertheless, it highlights the potential of corpus-based CDA in exploring multimodal and artistic texts, expanding the boundaries of discourse analysis.

10. Newspaper Discourses on the Ukraine War

"A qualitative corpus-assisted CDA compared coverage of the 2022 Russia–Ukraine war in "The Washington Post" and "The Moscow Times". Drawing on Systemic Functional Linguistics (SFL) and CDA, the study applied appraisal theory to analyze attitude resources (affect, judgment, appreciation). Findings revealed contrasting ideological stances: U.S. media framed Russia as the aggressor and Ukraine as the victim, while Russian media inverted these roles to emphasize Western provocation. This study illustrates the utility of integrating CL with SFL and CDA for analyzing evaluative language in conflict reporting. Its limitation lies in the small dataset, though it nonetheless provides insight into media polarization.

Comparative Analysis: Pakistani vs. International Studies

- **Corpus Resources**

International studies benefited from extensive, digitized corpora (e.g., millions of words from Xinhua and social media), while Pakistani studies often relied on smaller samples due to limited digitization. However,

Pakistani scholars compensated with mixed methods, integrating surveys and qualitative interviews, reflecting methodological adaptability.

- **Theoretical Application**

Pakistani research primarily drew on Fairclough and van Dijk, focusing on political and policy discourses. International research demonstrated more theoretical diversity, incorporating Wodak's DHA, SFL, and appraisal theory.

- **Thematic Breadth**

Both contexts addressed political and media discourses, but international studies also extended to health communication, online hate speech, and cultural texts. Pakistani studies remained focused on issues of immediate national relevance: Islamophobia, political rhetoric, education policy, and terrorism.

- **Methodological Strengths and Weaknesses**

International research excelled in large-scale corpus construction and computational analysis but risked reducing discourse to quantifiable patterns without sufficient qualitative depth. Pakistani research, while limited in corpus size, often integrated social context more explicitly, ensuring interpretative depth aligned with CDA's critical orientation.

Discussion

The synthesis of these studies highlights the complementary strengths of Pakistani and international scholarship. Pakistani research demonstrates CDA's relevance to contexts of linguistic diversity, political instability, and ideological struggle, though it urgently requires corpus development and digitization. International research showcases the power of corpus linguistics in handling large-scale data and detecting systematic ideological framing, but must guard against cultural reductionism and quantitative overemphasis. Together, these findings reaffirm that corpus-based CDA is most effective when it balances computational rigor with critical interpretation, integrates multiple theoretical frameworks, and situates discourse within its socio-political and historical contexts.

Findings and Implications

The meta-analysis of ten selected studies, five from Pakistan and five from international contexts, reveals several significant patterns in how corpus linguistics (CL) and Critical Discourse Analysis (CDA) are combined to study ideology, power, and representation. International studies benefit from large, digitized corpora, often exceeding millions of words. This enables robust quantitative analyses of

frequency, collocation, and concordance. By contrast, Pakistani research is constrained by limited access to digitized texts, resulting in smaller corpora and more reliance on qualitative or mixed methods.

Across contexts, Fairclough's (1995) three-dimensional model and van Dijk's (1997) socio-cognitive model remain dominant. International studies, however, exhibit more theoretical diversification by incorporating Wodak's Discourse-Historical Approach (2015) and systemic-functional linguistics (Halliday, 2014). Pakistani studies tend to focus more narrowly on Fairclough and van Dijk due to their direct applicability to political and media discourses.

Both Pakistani and international studies address political discourse, media representation, and ideological framing. International research, however, demonstrates broader thematic coverage, extending into health communication, hate speech on social media, and cultural products such as song lyrics. Pakistani research focuses primarily on urgent national concerns, such as Islamophobia, terrorism, education policy, and shifting political rhetoric.

International studies showcase computational sophistication but sometimes risk reducing discourse to statistical patterns without sufficient qualitative interpretation. Pakistani studies demonstrate strength in contextual interpretation but are limited in corpus size and technological integration.

Implications

This synthesis highlights the ongoing value of integrating corpus linguistics with CDA. The findings confirm that corpus-based CDA enhances "theoretical depth" by showing how ideological framing is embedded not only in isolated texts but also in broader discursive patterns. Importantly, the analysis shows that CDA frameworks remain flexible enough to address new types of discourse, including online hate speech and multimodal texts such as song lyrics.

The review underscores the need for "methodological pluralism. Corpus-based CDA is most effective when it balances quantitative rigor with qualitative depth. International research demonstrates the potential of large-scale corpora, while Pakistani research illustrates the importance of embedding analysis in socio-political realities. Future studies should integrate corpus methods with ethnographic or qualitative techniques to ensure both representativeness and contextual sensitivity.

Another implication is the importance of "corpus development" in underrepresented contexts. For Pakistan, there is an urgent need to build digitized corpora in Urdu and regional languages, which would allow for more inclusive analyses of ideology and power across multilingual landscapes.

Practically, corpus-based CDA offers insights for policymakers, educators, and media practitioners. For example, the analysis of education policy texts in Pakistan reveals how linguistic hierarchies marginalize regional languages, informing debates on curriculum reform. Similarly, studies of media

representation demonstrate how ideological framing shapes public perceptions of global crises, from terrorism to COVID-19.

Educationally, the findings suggest that corpus-based CDA can serve as a "pedagogical resource" in postgraduate programs. By combining computational tools with critical frameworks, students can develop both technical skills and critical awareness of language as a site of power. This approach can also help young researchers in Pakistan overcome limitations of corpus access by integrating smaller datasets with qualitative insights.

Ultimately, the meta-analysis shows that corpus-based CDA has significant potential for cross-cultural and comparative research. The ideological constructions of Islamophobia, terrorism, and war discourse in Pakistani and international contexts illustrate how discourses are locally situated yet globally interconnected. Future research should pursue "cross-corpus comparisons" across regions and languages to enhance generalizability and cultural inclusivity.

Conclusion

This article has presented a meta-analysis of corpus-based approaches to Critical Discourse Analysis (CDA) by synthesizing ten empirical studies, five from Pakistan and five from international contexts, published between 2015 and 2025. The analysis demonstrates that combining corpus linguistics (CL) with CDA provides a powerful methodological synergy that balances empirical rigor with critical interpretation. While CDA enables researchers to uncover hidden ideologies and structures of power in discourse, CL offers the tools to detect recurring linguistic patterns across large datasets, ensuring replicability and systematicity.

The findings reveal both convergences and divergences in Pakistani and international scholarship. Pakistani studies are strongly grounded in socio-political realities, addressing pressing issues such as Islamophobia, political rhetoric, terrorism, and educational language policies. They excel in contextual interpretation but are hindered by the limited availability of digitized corpora and the underrepresentation of regional languages. International studies, by contrast, benefit from extensive corpus resources and advanced computational methods, allowing for large-scale and statistically robust analyses of media, political, health, and cultural discourses. However, these studies sometimes risk privileging quantitative findings at the expense of deeper cultural or contextual insights.

The meta-analysis underscores the importance of methodological pluralism in discourse studies. Corpus-based CDA is most effective when it integrates computational techniques with critical, qualitative interpretation and situates findings within socio-political and historical contexts. For Pakistan, the development of open-access multilingual corpora represents a crucial step for advancing scholarship, while

internationally, greater attention to cultural sensitivity and qualitative depth will enhance the validity of findings.

In broader terms, the synthesis highlights the pedagogical value of corpus-based CDA. For postgraduate and doctoral programs, the integration of corpus tools with CDA frameworks equips students with both technical skills and critical awareness of language as a site of power and ideology. Additionally, the approach holds practical significance for policymakers, educators, and media practitioners, as it reveals how discourse contributes to shaping public perceptions and legitimizing social structures.

Ultimately, corpus-based CDA represents a promising and evolving paradigm in discourse studies. By bridging computational methods with critical inquiry, it not only deepens our understanding of how language reflects and constructs power but also provides pathways for more inclusive, context-sensitive, and globally relevant research in the years to come.

References

Abbas, F., & Shehzad, W. (2018). A critical discourse analysis of political discourse in Pakistan. "Pakistan Journal of Social Sciences, 38"(2), 471–484.

Baker, P. (2006). "Using corpora in discourse analysis. Continuum.

Baker, P. (2010). Sociolinguistics and corpus linguistics. "Edinburgh University Press".

Baker, P. (2012). Acceptable bias? Using corpus linguistics methods with critical discourse analysis. "Critical Discourse Studies, 9"(3), 247–256.

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.

Cheng, W. (2013). "Corpus-based approaches to critical discourse analysis. Bloomsbury.

Cooper, H. (2010). "Research synthesis and meta-analysis: A step-by-step approach" (4th ed.). Sage.

Fairclough, N. (1992). "Discourse and social change. Polity Press.

Fairclough, N. (1995). "Critical discourse analysis: The critical study of language. Longman.

Fairclough, N. (2001). "Language and power" (2nd ed.). Longman.

Fairclough, N. (2013). "Critical discourse analysis and language policy. Routledge.

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.

Halliday, M. A. K. (2014). "Halliday's introduction to functional grammar" (4th ed., revised by C. Matthiessen). Routledge.

Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), "Text and technology: In honour of John Sinclair" (pp. 157–176). John Benjamins.

Mautner, G. (2009). Corpora and critical discourse analysis. In P. Baker (Ed.), *Contemporary corpus linguistics" (pp. 32–46). Continuum.

McEnery, T., & Hardie, A. (2012). "Corpus linguistics: Method, theory and practice". Cambridge University Press.

Mushtaq, H., & Zahid, M. (2016). Critical discourse analysis of political TV talk shows. "Journal of Political Studies, 23"(1), 133–145.

Partington, A. (2010). Modern diachronic corpus-assisted discourse studies (MD-CADS) on UK newspapers: An overview of the project. "Corpora, 5"(2), 83–108.

Schegloff, E. A. (1997). Whose text? Whose context? "Discourse & Society, 8"(2), 165–187.

Scott, M. (1997). PC analysis of key words and key key words. "System, 25"(2), 233–245

Sinclair, J. (1991). "Corpus, concordance, collocation. Oxford University Press.

Stubbs, M. (1997). Whorf's children: Critical comments on critical discourse analysis. In A. Ryan & A. Wray (Eds.), "Evolving models of language" (pp. 100–116). Multilingual Matters.

Stubbs, M. (2001). "Words and phrases: Corpus studies of lexical semantics. Blackwell.

van Dijk, T. A. (1993). Principles of critical discourse analysis. "Discourse & Society, 4"(2), 249–283.

van Dijk, T. A. (1997). "Discourse as structure and process. Sage.

van Dijk, T. A. (2009). "Society and discourse: How social contexts influence text and talk. Cambridge University Press.

Vessey, R. (2017). Corpus approaches to discourse analysis: A critical review. "Discourse, Context & Media, 20", 1–10.

Widdowson, H. G. (1998). The theory and practice of critical discourse analysis. "Applied Linguistics, 19"(1), 136–151.

Wodak, R. (2001). "Methods of critical discourse analysis. Sage.

Wodak, R. (2015). "The politics of fear: What right-wing populist discourses mean. Sage.