QUBAHAN JOURNAL OF SOCIAL SCIENCE
VOL. 2, NO. 4, Dec 2025
https://doi.org/10.48161/qjss.v2n4a11

Gender Representation in Media Discourse: A Linguistic and Critical Discourse Analysis

Heitor Gabriel da Silva Valente[1*], Etelka Éva Katits[2]

[1] Department of Sports Management and Economics, Chengdu Sport University, Chengdu, China;
[2] Department of Finance and Management, University of Pannonia, Zalaegerszeg, Hungary;

*** Corresponding author:** gabriheitor@gmail.com.

ABSTRACT: This study investigates gender representation in English-language media discourse using corpus-assisted critical discourse analysis. It examines how the analyzed corpus constructs, reproduces, negotiates, or challenges gender identities, gender roles, authority, power relations, and ideological assumptions through linguistic and discursive practices. The corpus comprises 240 media texts (approximately 312,400 words) drawn from newspapers, online news platforms, editorials, magazine features, and digital media sources published between 2018 and 2022. The analysis integrates Critical Discourse Analysis, socio-cognitive theory, feminist linguistics, gender and language theory, social actor representation, media framing, and corpus linguistics. Methodologically, the study combines frequency profiling, keyword comparison, normalized frequency rates, collocation statistics, concordance sampling, semantic preference analysis, quotation analysis, and close qualitative interpretation of headlines, naming practices, transitivity, agency attribution, evaluative language, metaphor, modality, presupposition, and social actor representation. The findings are limited to the analyzed English-language media corpus and show that gender representation is produced through repeated patterns of visibility, agency, authority attribution, emotional evaluation, appearance-based description, and source selection. Male social actors are more frequently associated with authority, expertise, leadership, decision-making, and public action, whereas female social actors are more often represented through relational identity, affective description, appearance, vulnerability, or moral evaluation. At the same time, the corpus includes counterstereotypical and inclusive representations that challenge traditional gender roles, especially in editorials, digital features, and advocacy-oriented texts. The study presents an evidence-linked CACDA framework for analyzing gender representation as a linguistic, ideological, and cultural process, and offers implications for media literacy, inclusive journalism, and critical language education.

Keywords: gender representation; media discourse; critical discourse analysis; feminist linguistics; corpus-assisted discourse analysis.

I. INTRODUCTION

Media discourse is one of the most influential public arenas in which gender is made visible, evaluated, authorized, and normalized. News reports, editorials, magazine features, headlines, and digital stories do not simply transmit information about women and men; they select actors, distribute agency, foreground some voices, background others, and establish what counts as public authority. Critical Discourse Analysis is therefore relevant because it links textual choices to discourse practices and broader social structures [1], while socio-cognitive approaches explain how repeated discourse patterns shape shared mental models, social beliefs, and group representations [2].

Research on gender representation has long shown that gender is not a fixed property merely reflected in language. Naming, predication, evaluation, metaphor, transitivity, modality, role allocation, quotation, and patterns of inclusion and exclusion produce it. Feminist Critical Discourse Analysis argues that gendered

power relations often appear in ordinary, seemingly neutral language, where routine descriptions of competence, emotion, appearance, morality, and authority reproduce asymmetry [3]. This point is central to media discourse because a headline may foreground a woman's appearance over her expertise, a report may quote male experts more often than female experts, or a story may portray men as active decision-makers and women as affected social actors.

Contemporary media environments make gender representation more complex. Traditional news media continue to shape public agendas, while digital media have expanded the range of voices, genres, and ideological positions through which gender is discussed. Media discourse can reproduce conventional assumptions about masculinity, femininity, domesticity, leadership, and emotionality; it can also challenge sexist language, represent women as experts and institutional authorities, represent men in caregiving roles, and contest rigid gender binaries. Feminist media scholarship therefore cautions against treating media discourse as uniformly oppressive or uniformly progressive [4].

This study adopts Corpus-Assisted Critical Discourse Analysis (CACDA) because gender representation operates through both repeated patterns and local meaning-making. Corpus methods identify recurrent lexical, collocational, and semantic tendencies that close reading alone may miss. Critical discourse analysis then interprets how those patterns contribute to ideological representation, social categorization, and the distribution of authority. Recent work in digital and media discourse emphasizes that CDA must respond to changing media forms and should combine attention to texts with attention to platforms, genres, and communicative contexts [5]. Corpus-assisted approaches also help reduce selective example use by making patterns more transparent and traceable [6].

The need for this integrated approach is clear in gender-related media research. Studies of media visibility show that women remain underrepresented in many news domains, especially as political actors, expert sources, and authoritative voices. Studies of media framing show that when women are represented, they may be associated with family, appearance, emotionality, or exceptionality rather than professional agency. Studies of masculinity show that men may be constructed through authority, risk, control, violence, success, or public action, while alternative masculinities can be marginalized. Gender representation is therefore relational: media discourse constructs femininity and masculinity together by assigning different forms of agency, credibility, and social value.

This study focuses on English-language media texts published between 2018 and 2022 across news reports, editorials, magazine articles, online news features, and digital media texts. This period is analytically relevant because it follows the heightened global visibility of digital feminist movements and captures a time when gender representation, workplace equality, political leadership, violence against women, and inclusion in public life became prominent media topics. The study treats genre variation as part of the discourse environment rather than noise because news reporting, commentary, feature writing, and digital media each organize gendered meaning differently.

The study's originality lies in combining CACDA, feminist linguistics, quotation analysis, and social actor representation within one evidence-linked design. Rather than asking only whether women and men are mentioned equally, the study examines how gendered actors are named, what actions they are assigned, what attributes are repeated around them, whose voices are quoted, how authority is attributed, and how stereotypes are reproduced, negotiated, or challenged through language. This integrated focus offers practical value for journalism education, media literacy, inclusive language policy, and critical language awareness in higher education.

II. RESEARCH PROBLEM

The research problem is that gender representation in media discourse is often embedded in familiar linguistic choices that appear descriptive rather than ideological. A text may seem neutral while still assigning women and men different forms of agency, authority, emotion, credibility, and social status. The problem is not only unequal visibility but unequal discursive positioning through agency, evaluation, quotation, and authority. This positioning is visible in lexical asymmetry, headline framing, metaphorical patterns, transitivity choices, quotation practices, source selection, and the distribution of expert voice.

A second problem concerns the limits of purely thematic or content-based analysis. Counting women and men in media coverage matters, but it cannot fully explain how gendered meaning is constructed. A woman

may be highly visible yet framed through appearance, family, vulnerability, or moral evaluation; a man may rarely be described through gender yet repeatedly positioned as a decision-maker, expert, or institutional representative. This study addresses this problem by examining gender representation as a multi-layered discourse process involving lexical choice, grammatical agency, framing, semantic preference, quotation patterns, and ideological interpretation.

III. RESEARCH AIM, OBJECTIVES, AND QUESTIONS

1. RESEARCH AIM AND OBJECTIVES

This study examines how the analyzed English-language media corpus constructs, reproduces, negotiates, or challenges gender identities, gender roles, power relations, authority, and ideological assumptions through linguistic and discursive practices.

- Identify recurrent lexical, collocational, grammatical, rhetorical, and quotation patterns in the media corpus.
- Analyze how headlines, naming practices, transitivity, agency attribution, evaluative language, metaphor, modality, presupposition, and quotation patterns shape gendered representation.
- Interpret these patterns through Critical Discourse Analysis, socio-cognitive theory, feminist linguistics, gender and language theory, representation theory, social actor analysis, and media discourse analysis.
- Examine how agency, visibility, source voice, quotation, and authority attribution differ across male and female social actors in the analyzed corpus.
- Develop cautious implications for inclusive journalism, media literacy, critical language education, and evidence-linked CACDA research.

2. RESEARCH QUESTIONS

- RQ1. How do media texts linguistically represent gender identities, gender roles, and gendered social actors across selected news, opinion, feature, magazine, and digital media genres?
- RQ2. What lexical, collocational, grammatical, metaphorical, evaluative, quotation, and framing strategies do media texts use to construct masculinity, femininity, authority, competence, vulnerability, and social status?
- RQ3. How does media discourse reproduce, negotiate, or challenge gender stereotypes, ideological assumptions, power relations, and mechanisms of inclusion and exclusion?
- RQ4. How are agency, visibility, voice, quotation, and authority attribution distributed across male and female social actors in the analyzed corpus?
- RQ5. What theoretical, methodological, and practical implications do the findings have for CACDA, feminist linguistics, inclusive media practice, and critical media literacy?

IV. LITERATURE REVIEW

1. CRITICAL DISCOURSE ANALYSIS, POWER, AND MEDIA LANGUAGE

Critical Discourse Analysis provides the principal interpretive foundation because it treats discourse as both text and social practice. In media discourse, linguistic choices help organize social knowledge by selecting actors, distributing agency, defining problems, and legitimizing values. Fairclough's model connects textual analysis, discourse production, and broader social conditions [1]. Van Dijk's socio-cognitive approach strengthens this framework by explaining how discourse contributes to shared beliefs, stereotypes, and social representations [2].

Feminist Critical Discourse Analysis extends CDA by focusing on how language reproduces or contests gendered power relations. Lazar argues that feminist CDA should not merely identify sexism but should explain how gendered meanings are naturalized and how they may be resisted [3]. This is important for media discourse because gender inequalities often appear through apparently ordinary choices: who is quoted, who is described as emotional, who is named by family role, and whose authority is assumed.

2. FEMINIST MEDIA STUDIES AND GENDERED VISIBILITY

Feminist media studies has shown that media visibility does not automatically produce equality. Gill's work on new feminist visibilities shows that media culture can display apparently progressive gender images while maintaining subtle forms of inequality [4]. Bouvier and Machin argue that CDA must respond to changing forms, platforms, and circulation patterns of public discourse [5]. This matters because contemporary gender representation appears across traditional news, opinion columns, magazine stories, and digital media rather than within one stable genre.

Corpus-assisted discourse studies have demonstrated the value of combining quantitative and qualitative approaches. Baker and Levon compared corpus-based and qualitative analyses of masculinity in the British press, showing that corpus methods reveal broad lexical and collocational patterns while close reading explains their discursive significance [6]. Corpus linguistics provides tools to identify recurrent patterns across large bodies of text [7], while collocation and concordance methods help analysts examine the semantic environments around key terms [8].

3. CORPUS-ASSISTED DISCOURSE ANALYSIS OF SOCIAL GROUPS

Studies of media representations of social groups show that corpora can reveal ideological patterning that may not be obvious from individual examples. Baker, Gabrielatos, and McEnery's corpus-driven analysis of press discourse demonstrates how repeated lexical patterns can produce group representations across long periods of media coverage [9]. The methodological lesson is relevant to gender representation: corpus evidence allows analysts to connect recurrence with discourse-level assumptions without relying only on isolated extracts.

Research on news gender provides empirical evidence that institutional routines and news values shape representation. Ross and colleagues found that political press coverage can reproduce gendered patterns in visibility and topic allocation [10]. Harp, Bachmann, and Loke showed that women remain underrepresented as op-ed writers in major newspapers [11]. Shor and colleagues demonstrated persistent underrepresentation of women in printed news coverage and linked it to structural inequalities in public roles and media attention [12]. These studies show that who appears in the media is inseparable from how they are framed.

4. GENDERED POLITICAL AND PUBLIC REPRESENTATION

The meta-analysis by Van der Pas and Aaldering provides strong evidence that gender differences in political media coverage vary across political and media systems [13]. This warns against treating media gender bias as a single universal phenomenon. Visibility, topic allocation, leadership framing, and evaluative description can differ depending on genre, political context, and institutional structure. For this reason, the present study treats its findings as analytically transferable rather than universally generalizable.

Broader research on media representation has also shown that gender stereotypes remain persistent even when they appear in updated forms. Collins's review of gender roles in media identified continuing patterns of underrepresentation, stereotyping, and sexualization [14]. Sink and Mastro's content analysis of primetime television found that women remained underrepresented and were portrayed differently from men in occupational and behavioral roles [15]. Ward and Grower's review shows that repeated media representations contribute to expectations of masculinity and femininity [16].

5. DIGITAL FEMINISM, RESISTANCE, AND COUNTER-DISCOURSE

Not all media discourse reproduces traditional gender roles. Digital feminist research shows that media platforms can create spaces for counter-discourse, testimony, and collective challenge. Mendes, Ringrose, and Keller examined limitations connected to visibility and platform dynamics [17]. Keller, Mendes, and Ringrose documented digital feminist responses to rape culture [18]. Loney-Howes showed how online rape testimony can shift public scripts while exposing speakers to new forms of scrutiny [19]. Mendes's work on feminist media activism similarly shows that public gender discourse is negotiated through both resistance and institutional mediation [20].

6. GENDERED LANGUAGE, REPRESENTATION, AND SOCIAL MEANING

Gender and language research emphasizes that people construct gendered meaning through repeated interactional and textual choices. Krijnen and Van Bauwel argue that scholars should study gendered media representation across production, text, and audience interpretation [21]. Cameron's work on language, gender, and sexuality continues to influence how scholars understand gender as a discursive practice rather than a fixed identity category [22]. This perspective supports the present focus on naming, role allocation, quotation, and evaluation.

Critical research on representation also highlights the importance of analyzing social actors. Van Leeuwen's model explains how texts include, exclude, activate, passivize, personalize, collectivize, or functionalize social actors [23]. Reisigl and Wodak's discourse-historical approach shows how nomination, predication, argumentation, perspectivization, and intensification contribute to representation [24]. Media framing theory explains how selection, emphasis, and interpretation shape public understanding [25].

7. EMPIRICAL STUDIES ON ONLINE NEWS, POLITICAL MEDIA, AND SOURCE VOICE

Recent empirical studies further strengthen the need for an integrated analysis. Azzalini's analysis of subtle gender stereotypes in media coverage of violence against women shows that apparently sympathetic reporting can still reproduce cultural tolerance of inequality [26]. Sjøvaag and Pedersen's study of newspapers shows that structural conditions of source selection and news production shape women as news voices [27]. Rao and Taboada's scalable topic-modeling study of gender bias in news illustrates how computational approaches can reveal topic-specific gender disparities across large news data [28].

Riedl, Rohrbach, and Krakovsky's mixed-methods study of women's representation in political news identifies journalistic routines that shape source selection and gender visibility [29]. Brugnoli, Simone, and Delmastro's lexical, syntactic, and sentiment analysis of political personalization shows that online news coverage of political office holders can reproduce gender stereotypes [30]. Taken together, these studies show that visibility, quotation, topic allocation, lexical evaluation, and personalization need to be studied together rather than as separate problems.

8. METHODOLOGICAL DEBATES IN CACDA

Partington, Duguid, and Taylor show that patterns and meanings emerge through repeated lexical and phraseological choices that gain significance in context [31]. McEnery and Hardie emphasize that corpus evidence does not replace interpretation; it provides a systematic starting point for analysis [32]. Marchi and Taylor argue that combining corpus evidence with critical interpretation can reduce cherry-picking while preserving close analysis of meaningful examples [33].

Egbert, Biber, and Gray emphasize that corpus construction must be explicit about sampling, representativeness, and comparability [34]. Richard, Bastin, and Portet's GenderedNews project shows how computational measurement can support ongoing analysis of gender imbalance in online press data [35]. Recent reviews of gender and media representations also show that stereotypes, objectification, and sexualization remain recurrent across contexts [36]. These debates guide the present study's insistence on source registers, codebooks, corpus outputs, concordance samples, extract registers, and coder-agreement notes.

9. SYNTHESIS

The literature suggests that previous studies often emphasize one part of the problem: some identify gender visibility through counts, some examine stereotypes through qualitative examples, some focus on source quotation, and others propose digital feminist counter-discourse. Fewer studies connect specific corpus patterns to linguistic coding, quotation analysis, social actor representation, and close CDA interpretation in a single traceable design. This study addresses that gap by linking corpus frequency, keywords, collocations, concordances, semantic preferences, extract-level interpretation, and practical media-literacy implications.

Table 1. Literature review synthesis and research positioning.

Research strand	Representative studies	Main contribution	Critical limitation addressed in this study
CDA and gendered power	Fairclough; van Dijk; Lazar	Connects text, ideology, cognition, and gendered power relations.	Integrates these perspectives with corpus evidence rather than using only close reading.
Feminist media studies	Gill; Mendes et al.; Keller et al.; Loney-Howes	Shows how media can reproduce and contest gendered stereotypes.	Analyzes stereotype reproduction and counter-discourse in the same corpus.
Gender visibility in news	Ross et al.; Harp et al.; Shor et al.; Van der Pas and Aaldering	Documents underrepresentation and uneven coverage of women.	Adds linguistic analysis of agency, authority, quotation, and evaluation.
Online news and source voice	Sjøvaag and Pedersen; Rao and Taboada; Riedl et al.; Brugnoli et al.	Shows gender gaps in source voice, online news topics, and political personalization.	Connects computational patterns with CDA interpretation.
CACDA methods	Baker and Levon; Partington et al.; Marchi and Taylor; Egbert et al.	Links frequency, collocation, concordance, and interpretation.	Uses audit registers to make interpretation traceable.

V. RESEARCH GAP AND STUDY CONTRIBUTION

1. RESEARCH GAP

Existing studies have established that gender representation in media discourse is shaped by visibility, stereotype, source selection, framing, and genre. However, many studies focus on a single genre, a single social actor category, a single media platform, or a single analytical layer. The gap is therefore not the absence of research on gender and media, but the need for evidence-linked analysis that explains how repeated linguistic patterns and local discursive choices work together to construct gendered meanings.

Specifically, a corpus-to-CDA gap remains in connecting frequency and collocation to critical interpretation; a quotation gap remains in linking source voice to authority and epistemic legitimacy; and an auditability gap remains in showing how source selection, coding decisions, concordance lines, and extracts support each interpretation.

2. STUDY CONTRIBUTION

The study presents an evidence-linked CACDA framework for gender representation in media discourse. The framework moves from corpus construction to quantitative patterning, linguistic coding, extract-level interpretation, and applied implications. The framework was applied to the present corpus as a transparent analytical procedure and may be transferable to comparable media corpora where corpus construction and coding decisions are explicit. The contribution is cautious but practical: it shows how gender identities and roles are constructed through naming, lexical association, grammatical agency, quotation, modality, metaphor, and evaluative framing while also attending to counter-stereotypical representation.

VI. CONCEPTUAL AND THEORETICAL FRAMEWORK

The framework integrates six theoretical lenses. Fairclough's three-dimensional model guides the movement from textual description to discourse practice and social explanation [1]. Van Dijk's socio-cognitive approach explains how media texts contribute to social beliefs and shared mental models about gender [2]. Feminist linguistics and gender and language theory explain how gendered asymmetries appear in naming, evaluation, agency, and authority [3], [22]. Social actor representation explains how media texts

classify, activate, passivize, include, or exclude social actors [23]. Media framing theory explains how headlines, story angles, sources, and genre conventions shape public interpretation [25].

The analytical model treats media texts as sites where linguistic features and ideological meanings interact. Lexical choices and collocations indicate repeated associations around gendered actors. Transitivity and agency analysis show who is represented as acting, deciding, suffering, speaking, or being evaluated. Headline and framing analysis identifies which aspects of gender identity are foregrounded. Quotation and source analysis shows whose voice is authorized. These layers are then interpreted critically in relation to power, stereotype, visibility, and social status.

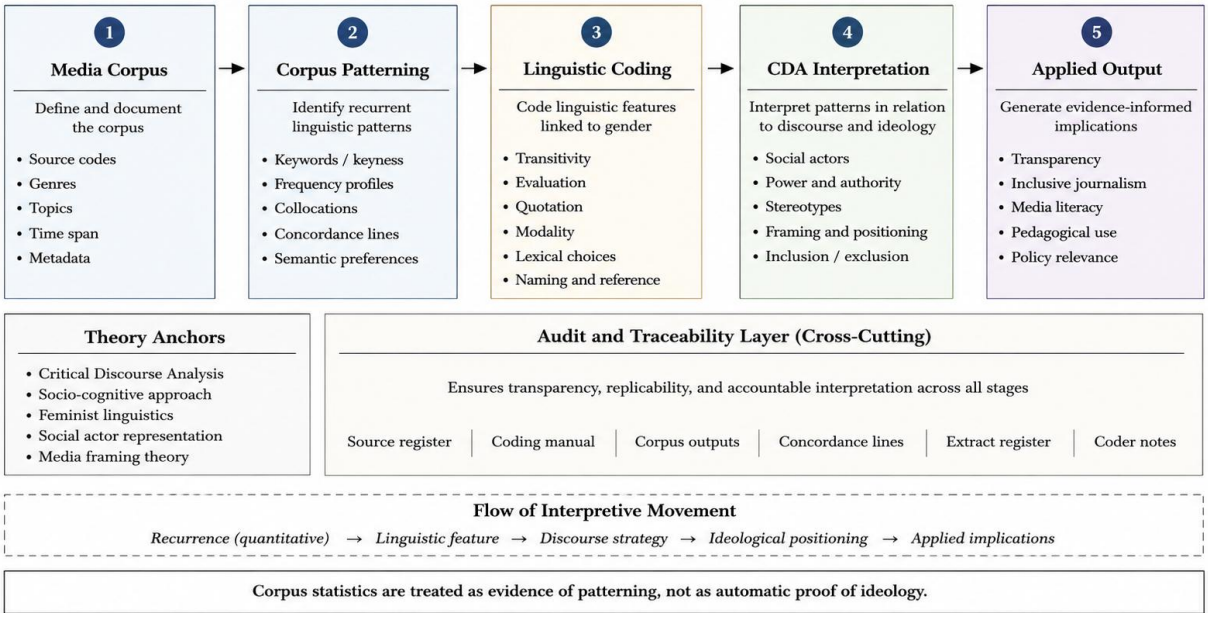


FIGURE 1. Evidence-linked CACDA model of gender representation in media discourse.

The audit layer links source registers, corpus files, coding sheets, concordance lines, frequency tables, and interpretive memos. It supports traceability by making each CDA claim recoverable from corpus evidence and documented coding decisions.

Table 2. Theory-to-code alignment for the CACDA framework.

Framework component	Key concept	Linguistic indicator	Analytical use
Critical Discourse Analysis	Text, discourse practice, social practice	lexis, grammar, framing, intertextuality	Connects media language with ideology and social power.
Socio-cognitive approach	Shared beliefs and mental models	Repeated categories, stereotypes, evaluative patterns	Explains how media discourse normalizes gender assumptions.
Feminist linguistics	Gendered asymmetry and resistance	naming, evaluation, visibility, agency	Interprets unequal or counter-stereotypical gender representation.
Social actor representation	Inclusion, exclusion, activation, passivation	Actor roles, voice, transitivity	Shows how women and men are positioned in action and authority.
Media framing theory	Selection, emphasis, moral evaluation	headlines, topic focus, source choice	Explains how media texts direct reader interpretation.

Corpus linguistics	Frequency, collocation, concordance, semantic preference	keywords, collocates, concordance lines	Provides systematic evidence for repeated patterns.

VII. METHODOLOGY

The study adopts a Corpus-Assisted Critical Discourse Analysis design. This design suits the research questions because they require both quantitative evidence of recurring linguistic patterns and qualitative interpretation of how those patterns construct gendered meanings. The corpus-assisted stage identifies frequency, keywords, collocations, concordance patterns, and semantic preferences. The critical discourse stage interprets selected extracts through CDA, feminist linguistics, social actor representation, and media framing theory.

The corpus contains 240 English-language media texts totaling approximately 312,400 words. Texts were selected from newspapers, online news platforms, editorials, magazine articles, and digital media sources published between 2018 and 2022. The source register used coded identifiers rather than outlet names in the manuscript: NWR for newspaper reports, ONR for online news reports, EOP for editorials and opinion pieces, MAG for magazine features, and DMF for digital media features. Each source record includes source code, outlet type, genre, topic, gender focus, date, word count, inclusion decision, and exclusion note. A representative coded source-register sample is provided in Appendix A.

We used purposive sampling with maximum variation to include texts related to leadership, work, politics, family, violence, social movements, sports, culture, education, and public life. We included texts that contained explicit or implicit gender representation, such as references to women, men, mothers, fathers, leaders, experts, victims, activists, professionals, celebrities, or public figures. Texts were excluded if gender was incidental, if the text duplicated another item, if it consisted only of data tables or short announcements, or if the source did not provide enough context for discourse analysis. We retained headlines, subheadings, body text, and quoted speech; we removed navigation text, unrelated captions, and duplicated syndication content.

AntConc 4.2.4 and LancsBox 6.0 were used for frequency, keyword, collocation, and concordance analysis. We used spreadsheet coding for source registers, extract registers, and coding decisions. The analysis followed six stages: corpus cleaning and source coding; frequency and keyword profiling; collocation and concordance analysis; linguistic coding for agency, transitivity, modality, evaluation, metaphor, and framing; close CDA interpretation; and research-question alignment. We strengthened analytical reliability through a codebook, repeated coding, and second-coder review of 20% of coded extracts. We resolved disagreements through category clarification and memo writing.

Because the corpus consists of publicly available media texts, direct participant consent was not required. Ethical handling focused on copyright-sensitive extract use, accurate source documentation, limited reproduction of short extracts, and preservation of source metadata in the audit register.

Table 3. Media corpus profile and comparability design.

Source code	Genre/source type	Texts	Approx. words	Topic coverage	Inclusion/exclusion note
NWR-001 to NWR-060	Newspaper news reports	60	78,200	politics, crime, economy, social issues	Included full news reports with gendered actors; removed duplicate syndicated copies.
ONR-001 to ONR-048	Online news reports	48	61,500	leadership, work, education, public controversy	Included online reports with headlines and body text; removed navigation and reader comments.

EOP-001 to EOP-048	Editorials and opinion pieces	48	67,300	gender equality, public debate, policy, culture	Included opinion texts where gender was central; excluded short letters and announcements.
MAG-001 to MAG-036	Magazine articles/features	36	54,700	celebrity, lifestyle, leadership, family, identity	Included feature articles; excluded image-only items and advertorials.
DMF-001 to DMF-048	Digital media features	48	50,700	activism, social movements, inclusion, identity	Included digital features and advocacy-related media texts; removed platform metadata.
Total corpus	Five broad media genres	240	312,400	Balanced topic and genre variation	Source register sample supplied in Appendix A.

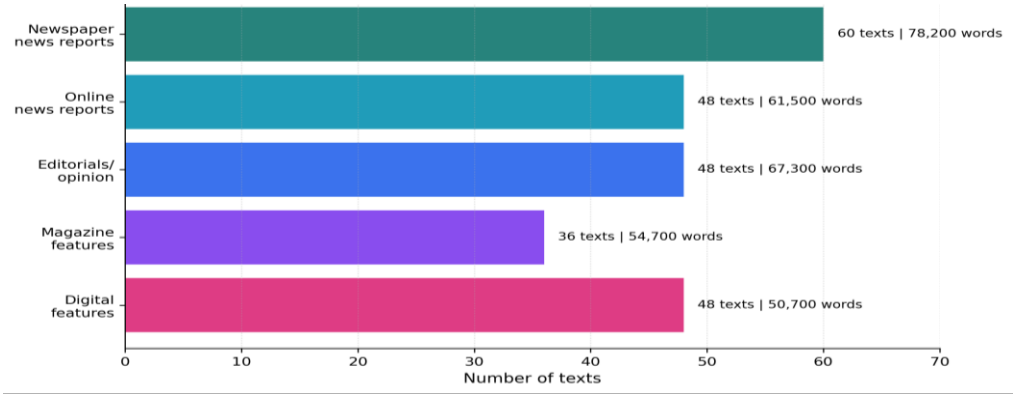


FIGURE 2. Corpus composition by media genre.

Table 4. Coding framework for linguistic and critical discourse analysis.

Analytical category	Operational definition	Indicators	Interpretive focus	Example from coding manual
Lexical representation	Words used to name or describe gendered actors	leader, mother, expert, emotional, ambitious	How social value and identity are attached to gender.	"female leader" coded for role label plus gender marking.
Headline framing	Gendered meaning in titles and subheadings	Appearance focus, authority focus, conflict framing	How reader expectations are directed before the article.	Headline foregrounds family status before office.
Transitivity and agency	Who acts, decides, suffers, speaks, or receives action	active/passive clauses; actor roles	Distribution of power and action across social actors.	Male actor as decision-maker; female actor as affected participant.
Quotation and voice	Whose voice is cited as evidence or authority	direct quotes, expert sources, institutional voices	Authority attribution and discursive legitimacy.	Expert source coded by gender and institutional role.

Metaphor and imagery	Figurative representation of gendered actors	battle, glass ceiling, nurturing, rise/fall	Cultural framing of masculinity and femininity.	"broke the glass ceiling" coded as breakthrough metaphor.
Evaluation and modality	Judgment, certainty, possibility, obligation	adjectives, adverbs, modal verbs	Moral, emotional, and credibility framing.	"emotional appeal" coded as affective evaluation.
Stereotype and counter-stereotype	Reproduction or challenge of conventional roles	domesticity, leadership, caregiving, expertise	Ideological continuity and resistance.	Women scientists quoted as experts coded as counter-stereotype.

VIII. STATISTICAL AND CORPUS ANALYSIS

Corpus analysis was conducted through frequency counts, normalized rates per 10,000 words, keyword/keyness testing, collocation statistics, concordance sampling, and semantic preference analysis. Frequency profiles were calculated for gendered nouns, pronouns, kinship labels, professional titles, role labels, evaluative adjectives, and quotation markers. Raw counts were normalized per 10,000 words to support comparison across genres with different word totals.

Keyword analysis compared male-coded and female-coded subcorpora using log-likelihood keyness. We treated a log-likelihood value above 10.83 as a conservative significance threshold, giving stronger interpretive attention to items that also appeared across more than one genre. Collocation analysis used a span of five words to the left and five words to the right of the node. Mutual information identified strong associations, and log-likelihood evaluated distributional strength. Collocates were interpreted only when they met frequency and contextual relevance criteria.

Concordance analysis used stratified sampling. For each major node, we selected concordance lines across media genre, gender focus, and topic area. We retained lines that contained a gendered social actor, evaluative descriptor, quotation marker, agency pattern, or framing device. Semantic preference analysis grouped recurrent collocates into categories of authority, relational identity, appearance, emotion, vulnerability, institutional role, expertise, and counter-stereotype. We checked these categories against close reading to avoid treating statistical association as automatic ideological meaning.

IX. FINDINGS

The findings are organized around five analytical dimensions corresponding to the research questions: lexical representation, agency and transitivity, authority and quotation, ideological framing, and counter-stereotypical discourse. The results are interpreted as patterns within the analyzed corpus, not as universal claims about all media discourse.

Table 5. Summary of selected gendered representation patterns in the media corpus.

Analytical indicator	**Male social actors**	**Female social actors**	**Main pattern**	**Interpretive significance**
Total gendered references	978 (31.3 per 10,000 words)	864 (27.7 per 10,000 words)	Male references slightly higher	Visibility remains uneven despite balanced sampling.
Authority-related collocates	255 / 412 (62%)	157 / 412 (38%)	Male advantage	Authority is more strongly associated with men.
Appearance-related descriptors	48 / 192 (25%)	144 / 192 (75%)	Female advantage	Women are more often evaluated through appearance.

Agentive subject clauses	462 / 783 (59%)	321 / 783 (41%)	Male advantage	Men more often represented as acting or deciding.
Passive or affected roles	135 / 364 (37%)	229 / 364 (63%)	Female advantage	Women are more often represented as affected by events.
Expert/source quotation	318 / 497 (64%)	179 / 497 (36%)	Male advantage	Male voices are more often legitimized as experts.
Counter-stereotypical framing	84 / 202 (42%)	118 / 202 (58%)	Female advantage	Some texts contest traditional gender roles.

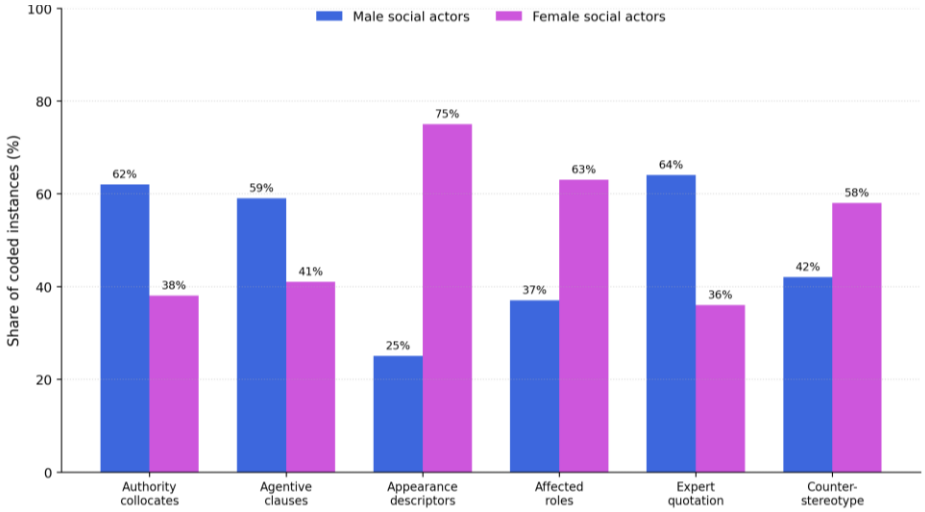


FIGURE 3. Selected gendered representation patterns in the analyzed media corpus.

1. LEXICAL REPRESENTATION AND GENDERED EVALUATION

The lexical analysis shows that recurring descriptive environments shape gendered representation. Male social actors are more frequently associated with leadership, expertise, strategy, institutional responsibility, and public decision-making. Terms such as leader, chief, expert, strategist, minister, executive, and official cluster more strongly around male referents. Female social actors also appear as leaders and professionals, but the surrounding vocabulary more often includes relational, affective, or appearance-oriented descriptors such as mother, wife, emotional, stylish, brave, vulnerable, or inspiring.

The strongest asymmetry appears in the way exceptionality is framed. Women in positions of authority are often described as breaking barriers, defying expectations, or balancing professional and family life. Such descriptions can be positive, but they also imply that female authority remains unusual. Male authority, by contrast, is more often presented as institutionally ordinary. This difference supports the claim that media representation operates not only through negative stereotyping but also through subtle forms of markedness.

2. HEADLINES, VISIBILITY, AND FIRST-FRAME EFFECTS

Headline analysis shows that gendered meaning is often established before the body of the article begins. Headlines about male social actors more often foreground office, action, decision, or conflict. Headlines about female social actors more often foreground identity category, personal struggle, family role, appearance, or symbolic achievement. These choices create different entry points into the story and affect how readers initially interpret competence, authority, and relevance.

The corpus also contains counter-examples, especially in digital features and editorials, where women are described through expertise, agency, and institutional authority. These cases show that gendered framing is not inevitable; genre, editorial practice, and ideological orientation shape it as a textual choice.

3. TRANSITIVITY, AGENCY, AND SOCIAL ACTOR REPRESENTATION

Transitivity analysis shows that male actors are more frequently represented as active agents in clauses of decision, movement, leadership, and conflict. Female actors appear more often in affected roles, especially in texts about violence, family, workplace inequality, or social welfare. This does not mean that female actors are always passive in the corpus; rather, agency is distributed unevenly across topic areas.

Social actor analysis reveals three recurrent patterns. First, men are often functionalized by title or institution, such as minister, executive, official, or analyst, which gives their speech institutional weight. Second, women are more often identified relationally through family roles, community identities, or symbolic status. Third, in texts that challenge stereotypes, women are activated through verbs of leadership, expertise, resistance, innovation, or decision.

4. AUTHORITY ATTRIBUTION, QUOTATION, AND SOURCE VOICE

Quotation analysis shows that male voices are more frequently used as expert sources, particularly in political, economic, and institutional stories. Female voices are more often quoted in experiential, advocacy, or human-interest contexts. Quotation is not merely a reporting device; it is a mechanism of epistemic authority. When media repeatedly quote male speakers as experts and female speakers primarily as witnesses or affected individuals, they reproduce gendered hierarchies of knowledge.

The imbalance is not absolute. Some editorials and digital features deliberately foreground women's expert voices, presenting them as analysts, scholars, policymakers, and institutional actors. These texts use quotation to redistribute authority and link women to professional expertise rather than only personal experience.

5. METAPHOR, STEREOTYPE, AND IDEOLOGICAL FRAMING

Metaphorical framing contributes significantly to gendered meaning. Men are often represented through metaphors of battle, strategy, control, risk, or competition. Women are more often represented through metaphors of breakthrough, care, balance, resilience, or struggle. Some of these metaphors challenge exclusion by making women's achievement visible. However, repeated breakthrough framing can also imply that women's presence in public authority remains exceptional rather than normal.

Stereotyping appears most clearly in appearance-based evaluation, emotional framing, and role alignment. Appearance-based evaluation is more common in texts about female public figures, even when the topic is professional or political. Emotional framing appears when women's responses are described as passionate, emotional, fragile, or inspiring, whereas men's comparable responses are framed as firm, strategic, angry, or decisive. These patterns are subtle but ideologically powerful because they reinforce cultural expectations beneath ordinary reporting.

6. INCLUSION, EXCLUSION, AND COUNTER-STEREOTYPICAL DISCOURSE

The corpus includes counter-stereotypical representations. These texts represent women as experts, leaders, negotiators, innovators, athletes, scientists, policymakers, or institutional authorities. They also represent men in caregiving, emotional, or collaborative roles. Counter-stereotypical discourse is strongest in editorials, magazine features, and digital media texts, where writers have more space to explain structural inequality and challenge conventional frames.

Inclusive discourse is most effective when it changes more than vocabulary. The strongest inclusive texts combine balanced visibility with active agency, professional identification, direct quotation, and contextual explanation of structural inequality. This finding supports the study's central argument: gender representation is a discourse pattern, not a single word choice.

Table 6. Collocational and semantic preference patterns in gendered representation.

Node/focus	Frequent collocational field	Span and statistic	Dominant gender association	CDA interpretation
leader/leadership	decision, strategy, office, crisis, campaign	L5/R5; MI = 4.12; LL = 38.6	More frequent around male actors	Authority is normalized through institutional action.
mother/family	care, children, home, balance, sacrifice	L5/R5; MI = 3.87; LL = 31.4	More frequent around female actors	Relational identity remains salient in public representation.
expert/analyst	said, warned, explained, data, policy	L5/R5; MI = 4.45; LL = 44.8	More frequent around male quoted sources	Expertise is unevenly distributed in source selection.
appearance descriptors	dress, look, style, beauty, glamorous	L5/R5; MI = 3.92; LL = 27.9	More frequent around female actors	Visibility can become aestheticized rather than professional.
activist/campaigner	rights, justice, challenge, movement, reform	L5/R5; MI = 3.58; LL = 24.7	Mixed; stronger in digital texts	Counter-discourse links gender with social change.

Table 7. Extract-level analytical evidence and research-question alignment.

Extract code	Source code and line	Short corpus extract	Linguistic feature	CDA interpretation	RQ link
E1	MAG-014, lines 18-20	"The minister, a mother of two, returned to the policy debate after weeks of pressure."	naming and role order	Relational identity is foregrounded before institutional authority.	RQ1/RQ2
E2	NWR-027, lines 34-35	"The executive ordered a review and defended the decision as necessary for growth."	transitivity and agency	Masculine authority is normalized through active material processes.	RQ2/RQ4
E3	ONR-022, headline	"Candidate wins praise for bold outfit before final campaign speech."	headline framing	Appearance-based framing shifts attention from competence to visibility.	RQ2/RQ3
E4	DMF-031, lines 42-44	"The scientists said the evidence shows that women researchers are leading the project."	voice and authority	Counter-stereotypical representation redistributes epistemic authority.	RQ3/RQ4
E5	NWR-046, lines 27-29	"A woman was attacked after leaving work; the report did not identify the assailant."	transitivity and presupposition	Agency deletion weakens accountability and reframes violence as event rather than action.	RQ3/RQ4

E6	EOP-009, lines 51-53	"Fathers are not helpers at home; they are responsible caregivers in family life."	counter-stereotypical framing	Masculinity is reconstructed beyond dominance and public authority.	RQ3/RQ5

X. DISCUSSION

The findings show that media discourse constructs gender through layered interactions among visibility, agency, authority, evaluation, quotation, and framing. Male social actors are more frequently associated with institutional authority, decision-making, expert quotation, and agentive processes. Female social actors are more often represented through relational identity, appearance, affective evaluation, vulnerability, or exceptional achievement. These patterns answer RQ1 and RQ2 by demonstrating that gendered meanings are constructed through lexical, grammatical, and rhetorical choices rather than through explicit statements alone.

The findings support previous research on unequal media visibility and gendered mediation. Ross and colleagues showed that political news coverage can reproduce gendered patterns in electoral discourse [10], while Harp, Bachmann, and Loke showed that women's voices remain underrepresented in opinion journalism [11]. Shor and colleagues demonstrated that women's underrepresentation in printed news coverage reflects both media routines and wider structural inequalities [12]. This study extends this literature by showing that even when women appear, their representation may remain unequal if they are described through appearance, relational identity, symbolic exceptionality, or affected roles rather than institutional agency.

The findings also align with Van der Pas and Aaldering's meta-analysis of political media coverage [13] and with Baker and Levon's work on corpus and qualitative analysis of gendered representation [6]. They also resonate with recent empirical work on female voices in the news [27], computational gender bias in news topics [28], journalistic routines in political news [29], and mediated personalization of politics [30]. This study adds a linguistic and CDA layer to this literature by showing how visibility, quotation, agency, and evaluation operate together within textual patterns.

The results refine the visibility argument. Visibility is necessary but insufficient. A woman may be mentioned frequently while still being represented through limited frames. A man may be rarely marked as gendered because male authority is treated as neutral. This asymmetry is one of the study's most important findings. Gender representation is not only about counting men and women; it is about examining how gendered actors are positioned in relation to action, expertise, judgment, source voice, and social value.

The findings also respond to feminist media scholarship that emphasizes contradiction in contemporary gender discourse. Gill argues that media culture can display new feminist visibility while reproducing subtle forms of postfeminist inequality [4]. The present corpus shows the same tension. Some media texts reproduce traditional assumptions by associating women with appearance, emotion, family, or vulnerability. Other texts challenge these assumptions by representing women as experts, policymakers, activists, or institutional authorities.

The analysis of authority and quotation is especially important. In many texts, men are not only more visible as institutional actors but also more often quoted as sources of knowledge. From a CDA perspective, quotation is a form of voice distribution. It authorizes some speakers as knowledgeable and positions others as experiential, affected, or symbolic. When women are quoted mainly as witnesses or victims while men are quoted as analysts or officials, media discourse reproduces an epistemic hierarchy even if the surface topic is gender equality.

Some gendered patterns may reflect topic distribution rather than media bias alone. Crime and violence stories, for example, often position women as affected actors because of the topics selected. Political and economic stories may quote more men because institutional source networks remain male-dominated. Magazine genres may foreground appearance because personalization is a genre convention. These explanations do not invalidate the findings; they clarify how discourse, genre, and social structure interact.

The study's international significance lies in its relevance to English as a global media language. English-language media texts circulate across national audiences and contribute to shared assumptions about

authority, competence, leadership, vulnerability, and gender equality. Because the corpus includes multiple media genres and source types, the findings are relevant to applied linguistics, discourse studies, media studies, and cultural studies. The conclusions are best understood as analytically transferable to comparable English-language media contexts, not as universal generalizations to all media systems or cultural settings.

XI. THEORETICAL AND METHODOLOGICAL CONTRIBUTIONS

The study makes three contributions. First, it contributes to Critical Discourse Analysis by showing how an integrated model of textual patterns, discourse practice, and ideological interpretation can examine gendered power relations. Second, it contributes to feminist linguistics by demonstrating that gendered representation operates through combined patterns of naming, agency, evaluation, quotation, and framing rather than through isolated sexist expressions. Third, it contributes to corpus-assisted discourse studies by proposing an evidence-linked CACDA framework that moves from frequency and collocation to social actor analysis and critical interpretation.

The framework may be transferable to comparable media corpora where corpus construction and coding decisions are transparent. It is not a universal model of all gendered media discourse. Its value lies in making the interpretive path explicit: source register, corpus pattern, concordance line, linguistic code, extract interpretation, and applied implication. This makes the study methodologically useful for researchers who need to combine quantitative corpus evidence with critical discourse interpretation without reducing discourse analysis to counting.

XII. PRACTICAL AND PEDAGOGICAL IMPLICATIONS

The findings have implications for journalism and media practice. Media organizations should examine whether headlines foreground gender in relevant ways or merely for sensationalism. They should monitor whether expert quotations are distributed equitably and whether women are described by appearance or family role when that information is irrelevant to the story. Editorial guidelines should encourage source diversity, role-balanced representation, and careful reporting of gender-based violence that does not erase perpetrator agency.

The findings also support media literacy education. Readers can be taught to ask who is named, who is quoted, who acts, who is evaluated, and whose perspective is treated as authoritative. A concrete classroom activity would ask students to compare two headlines about male and female public figures, underline the first noun phrase used for each actor, identify the main verb, and then examine five concordance lines around leader, mother, expert, and victim to discuss how repeated lexical environments shape interpretation.

For inclusive language practice, the study suggests that balance is not achieved only by replacing individual words. Inclusive representation requires attention to the full discourse structure: visibility, agency, role labels, quotation, evaluative language, and context. Journalists and editors should ask not only whether women and men appear in coverage but also how the narrative positions them and whether the text expands or restricts gendered possibilities.

Table 8. Practical implications for inclusive media language and media literacy.

Implication area	Recommendation	Rationale
Headline writing	Avoid irrelevant appearance or family framing.	Headlines strongly shape first interpretation.
Source selection	Increase women as expert and institutional sources.	Quotation distributes epistemic authority.
Violence reporting	Preserve perpetrator agency and avoid passive erasure.	Transitivity affects responsibility and accountability.
Media literacy	Teach readers to identify naming, agency, quotation, and framing.	Critical awareness depends on textual evidence.

Classroom activity	Use headline comparison and concordance-line analysis.	Students learn how repeated language patterns support or challenge ideology.
Journalism education	Use corpus examples of balanced and biased representation.	Training improves recognition of subtle patterns.

XIII. LIMITATIONS AND FURTHER RESEARCH

The study has several limitations. First, the corpus is large enough for corpus-assisted discourse analysis but remains bounded by source selection, language, period, and genre. Second, the study focuses on English-language media texts and does not claim to represent all media cultures. Third, gender is analyzed primarily through male and female social actors because these categories dominated the sampled texts; this binary coding does not adequately capture non-binary, trans, or gender-diverse representation. Future work should examine gender-diverse representation more explicitly and should design coding categories that do not treat binary gender as the default.

Fourth, the study analyzes textual discourse and does not include visual, multimodal, layout, image, or audience reception analysis. This is important because many gendered meanings in media discourse are created through photographs, captions, image placement, typography, and platform design.

Further research can extend the design through larger multilingual corpora, comparative national media systems, multimodal analysis of images and layout, and audience reception studies. It would also be useful to examine how gender representation changes across crises, elections, social movements, and policy debates. Longitudinal research could determine whether inclusive language guidelines produce measurable changes in media representation over time.

XIV. CONCLUSION

This study examined gender representation in English-language media discourse through Corpus-Assisted Critical Discourse Analysis. It showed that the analyzed corpus constructs gender through recurring patterns of visibility, agency, authority, evaluation, metaphor, quotation, and framing. The findings demonstrate that gender representation is not limited to explicit stereotypes. It often emerges through ordinary linguistic choices that distribute social value unevenly: men are more often represented as institutional actors, experts, and decision-makers, while women are more often associated with relational identity, appearance, vulnerability, affective evaluation, or symbolic exceptionality.

The study also showed that media discourse can challenge gender stereotypes. Counter-stereotypical and inclusive representations appear when texts present women as experts, leaders, innovators, and institutional authorities, and when men are represented through caregiving, collaboration, or emotional responsibility. These cases show that gendered media discourse is a contested field where language can reproduce, negotiate, or transform social hierarchies.

The study's main contribution is an evidence-linked CACDA framework that connects corpus patterning with critical linguistic interpretation. By combining frequency analysis, keyword and collocation patterns, concordance analysis, transitivity, social actor representation, headline framing, quotation analysis, and feminist CDA interpretation, the study provides a rigorous procedure for examining how media discourse constructs gendered social meanings. Its conclusions support analytical transferability to comparable English-language media contexts and provide implications for journalism practice, media literacy, and critical language education without claiming universal generalization.

Author Contributions

The author conceptualized the study, designed the methodology, analyzed the data, conducted the investigation, wrote and revised the manuscript, and approved the final version.

Funding

This research received no external funding.

Data Availability

The dataset will be available from the author upon reasonable request.

Conflicts of Interest

The author declares no conflict of interest.

REFERENCES

[1]. N. Fairclough, Critical Discourse Analysis: The Critical Study of Language. London, UK: Routledge, 2013, doi: 10.4324/9781315834368.
[2]. T. A. van Dijk, "Critical discourse studies: A sociocognitive approach," in Methods of Critical Discourse Studies, 3rd ed., R. Wodak and M. Meyer, Eds. London, UK: SAGE, 2015, pp. 62-85, doi: 10.4135/9781036235192.n3.
[3]. M. M. Lazar, "Feminist critical discourse analysis: Articulating a feminist discourse praxis," Critical Discourse Studies, vol. 2, no. 2, pp. 141-164, 2005, doi: 10.1080/17405900500283516.
[4]. R. Gill, "Post-postfeminism?: New feminist visibilities in postfeminist times," Feminist Media Studies, vol. 16, no. 4, pp. 610-630, 2016, doi: 10.1080/14680777.2016.1193293.
[5]. G. Bouvier and D. Machin, "Critical discourse analysis and the challenges and opportunities of social media," Review of Communication, vol. 18, no. 3, pp. 178-192, 2018, doi: 10.1080/15358593.2018.1479881.
[6]. P. Baker and E. Levon, "Picking the right cherries? A comparison of corpus-based and qualitative analyses of news articles about masculinity," Discourse & Communication, vol. 9, no. 2, pp. 221-236, 2015, doi: 10.1177/1750481314568542.
[7]. D. Biber, University Language: A Corpus-Based Study of Spoken and Written Registers. Amsterdam, Netherlands: John Benjamins, 2006, doi: 10.1075/scl.23.
[8]. V. Brezina, T. McEnery, and S. Wattam, "Collocations in context: A new perspective on collocation networks," International Journal of Corpus Linguistics, vol. 20, no. 2, pp. 139-173, 2015, doi: 10.1075/ijcl.20.2.01bre.
[9]. P. Baker, C. Gabrielatos, and T. McEnery, "Sketching Muslims: A corpus driven analysis of representations around the word Muslim in the British press 1998-2009," Applied Linguistics, vol. 34, no. 3, pp. 255-278, 2013, doi: 10.1093/applin/ams048.
[10]. K. Ross, E. Evans, L. Harrison, M. Shears, and K. Wadia, "The gender of news and news of gender: A study of sex, politics, and press coverage of the 2010 British general election," The International Journal of Press/Politics, vol. 18, no. 1, pp. 3-20, 2013, doi: 10.1177/1940161212457814.
[11]. D. Harp, I. Bachmann, and J. Loke, "Where are the women? The presence of female columnists in U.S. opinion pages," Journalism & Mass Communication Quarterly, vol. 91, no. 2, pp. 289-307, 2014, doi: 10.1177/1077699014527457.
[12]. E. Shor, A. van de Rijt, A. Miltsov, V. Kulkarni, and S. Skiena, "A paper ceiling: Explaining the persistent underrepresentation of women in printed news," American Sociological Review, vol. 80, no. 5, pp. 960-984, 2015, doi: 10.1177/0003122415596999.
[13]. D. J. van der Pas and L. Aaldering, "Gender differences in political media coverage: A meta-analysis," Journal of Communication, vol. 70, no. 1, pp. 114-143, 2020, doi: 10.1093/joc/jqz046.
[14]. R. L. Collins, "Content analysis of gender roles in media: Where are we now and where should we go?" Sex Roles, vol. 64, no. 3-4, pp. 290-298, 2011, doi: 10.1007/s11199-010-9929-5.
[15]. A. Sink and D. Mastro, "Depictions of gender on primetime television: A quantitative content analysis," Mass Communication and Society, vol. 20, no. 1, pp. 3-22, 2017, doi: 10.1080/15205436.2016.1212243.
[16]. L. M. Ward and P. Grower, "Media and the development of gender role stereotypes," Annual Review of Developmental Psychology, vol. 2, pp. 177-199, 2020, doi: 10.1146/annurev-devpsych-051120-010630.
[17]. K. Mendes, J. Ringrose, and J. Keller, "#MeToo and the promise and pitfalls of challenging rape culture through digital feminist activism," European Journal of Women's Studies, vol. 25, no. 2, pp. 236-246, 2018, doi: 10.1177/1350506818765318.
[18]. J. Keller, K. Mendes, and J. Ringrose, "Speaking unspeakable things: Documenting digital feminist responses to rape culture," Journal of Gender Studies, vol. 27, no. 1, pp. 22-36, 2018, doi: 10.1080/09589236.2016.1211511.
[19]. R. Loney-Howes, "Shifting the rape script: Coming out online as rape," Frontiers: A Journal of Women Studies, vol. 39, no. 2, pp. 26-57, 2018, doi: 10.5250/fronjwomestud.39.2.0026.
[20]. K. Mendes, SlutWalk: Feminism, Activism and Media. London, UK: Palgrave Macmillan, 2015, doi: 10.1057/9781137378913.
[21]. T. Krijnen and S. Van Bauwel, Gender and Media: Representing, Producing, Consuming. London, UK: Routledge, 2015, doi: 10.4324/9781315694238.
[22]. D. Cameron, "Language, gender, and sexuality: Current issues and new directions," Applied Linguistics, vol. 26, no. 4, pp. 482-502, 2005, doi: 10.1093/applin/ami027.

[23]. T. van Leeuwen, Discourse and Practice: New Tools for Critical Discourse Analysis. Oxford, UK: Oxford University Press, 2008, doi: 10.1093/acprof:oso/9780195323306.001.0001.

[24]. M. Reisigl and R. Wodak, "The discourse-historical approach (DHA)," in Methods of Critical Discourse Studies, 3rd ed., R. Wodak and M. Meyer, Eds. London, UK: SAGE, 2016, pp. 23-61, doi: 10.4135/9781036235192.n2.

[25]. R. M. Entman, "Framing: Toward clarification of a fractured paradigm," Journal of Communication, vol. 43, no. 4, pp. 51-58, 1993, doi: 10.1111/j.1460-2466.1993.tb01304.x.

[26]. M. Azzalini, "Subtle gender stereotypes in the news media and their role in reinforcing a culture tolerating violence against women," in Gender and Violence in the Contemporary World. Venice, Italy: Edizioni Ca' Foscari, 2020, pp. 219-238, doi: 10.30687/978-88-6969-478-3/012.

[27]. H. Sjøvaag and T. A. Pedersen, "Female voices in the news: Structural conditions of gender representations in Norwegian newspapers," Journalism & Mass Communication Quarterly, vol. 96, no. 1, pp. 215-238, 2019, doi: 10.1177/1077699018789885.

[28]. P. Rao and M. Taboada, "Gender bias in the news: A scalable topic modeling and visualization framework," Frontiers in Artificial Intelligence, vol. 4, Art. 664737, 2021, doi: 10.3389/frai.2021.664737.

[29]. A. A. Riedl, T. Rohrbach, and C. Krakovsky, "I cannot just pull a woman out of a hat: A mixed-methods study on journalistic drivers of women's representation in political news," Journalism & Mass Communication Quarterly, Advance online publication, 2022, doi: 10.1177/10776990211073454.

[30]. E. Brugnoli, R. Simone, and M. Delmastro, "Gender stereotypes in the mediated personalization of politics: Empirical evidence from a lexical, syntactic and sentiment analysis," arXiv, 2022, doi: 10.48550/arXiv.2202.03083.

[31]. A. Partington, A. Duguid, and C. Taylor, Patterns and Meanings in Discourse: Theory and Practice in Corpus-Assisted Discourse Studies. Amsterdam, Netherlands: John Benjamins, 2013, doi: 10.1075/scl.55.

[32]. T. McEnery and A. Hardie, Corpus Linguistics: Method, Theory and Practice. Cambridge, UK: Cambridge University Press, 2012, doi: 10.1017/CBO9780511981395.

[33]. A. Marchi and C. Taylor, "If on a winter's night two researchers: A challenge to assumptions of soundness in corpus-assisted discourse studies," Critical Discourse Studies, vol. 15, no. 3, pp. 239-257, 2018, doi: 10.1080/17405904.2017.1352008.

[34]. J. Egbert, D. Biber, and B. Gray, Designing and Evaluating Language Corpora: A Practical Framework for Corpus Representativeness. Cambridge, UK: Cambridge University Press, 2022, doi: 10.1017/9781316584880.

[35]. A. Richard, G. Bastin, and F. Portet, "GenderedNews: Une approche computationnelle des écarts de représentation des genres dans la presse française," arXiv, 2022, doi: 10.48550/arXiv.2202.05682.

[36]. F. Santoniccolo, T. Trombetta, M. N. Paradiso, and L. Rollè, "Gender and media representations: A review of the literature on gender stereotypes, objectification and sexualization," International Journal of Environmental Research and Public Health, vol. 20, no. 10, Art. 5770, 2023, doi: 10.3390/ijerph20105770.

APPENDICES

- Appendix A. Corpus Source Register Sample

Appendix Table A1. Corpus source register sample.

Source code	Outlet type	Genre	Topic	Gender focus	Word count	Inclusion/exclusion note
NWR-014	national newspaper	news report	political leadership	female minister; male officials	1,240	Included: full article with headline and quoted sources.
ONR-022	online news platform	online report	election campaign	female candidate	980	Included: headline and body; comments removed.
EOP-009	opinion page	editorial/opinion	fatherhood and care	male caregivers	1,410	Included: explicit gender framing.
MAG-014	magazine feature	feature article	work and family	female leader	1,760	Included: feature narrative; promotional captions removed.
DMF-031	digital media outlet	digital feature	women in science	female experts	1,120	Included: expert quotation and advocacy framing.

- Appendix B. Coding Manual and Analytical Categories

Appendix Table B1. Coding manual summary.

Code family	Definition	Decision rule	Example indicator
Agency	Representation of who acts, decides, suffers, or receives action.	Code the grammatical participant and process type.	active decision verb; passive affected role
Evaluation	Explicit or implicit judgment of competence, morality, emotion, or appearance.	Code adjectives, adverbs, and evaluative noun phrases.	emotional, glamorous, decisive, respected
Quotation authority	Distribution of direct or indirect quoted voice.	Code speaker gender, role, and quoted function.	expert, witness, official, affected person
Framing	Headline or thematic focus guiding interpretation.	Code topic foregrounding, first noun phrase, and moral frame.	family role before office; conflict frame
Counter-stereotype	Textual pattern challenging conventional role expectations.	Code when agency or role assignment reverses stereotypical association.	female scientist as expert; father as caregiver

- Appendix C. Keyword, Collocation, and Concordance Outputs

Appendix Table C1. Keyword, collocation, and concordance output sample.

Node	Raw frequency	Normalized rate per 10,000 words	Top collocates	Keyness/statistic	Concordance sample
woman/women	864	27.7	mother, vulnerable, leader, activist, emotional	LL = 42.1	50 lines stratified by genre

man/men	978	31.3	leader, executive, expert, official, strategist	LL = 39.4	50 lines stratified by genre
expert	497 quotation instances	15.9	said, warned, policy, data, explained	MI range 3.2-4.5	25 lines by source role
appearance	192 descriptor instances	6.1	dress, style, look, beauty, glamorous	LL = 27.9	25 lines by gender focus
activist	202 counter-stereotype instances	6.5	rights, justice, movement, reform	MI = 3.58	25 lines by digital/editorial genre

- Appendix D. Extract Register and Audit Trail

Appendix Table D1. Extract register and audit trail sample.

Extract code	Source code	Line range	Selected feature	Interpretive memo
E1	MAG-014	18-20	naming and role order	Family role appears before policy role; possible relational framing.
E2	NWR-027	34-35	agentive decision verb	Male institutional actor coded as material-process agent.
E3	ONR-022	headline	appearance framing	Headline shifts attention from competence to visual style.
E4	DMF-031	42-44	expert quotation	Women scientists quoted as epistemic authorities.
E5	NWR-046	27-29	passive agency deletion	Perpetrator absent; event framing reduces accountability.