

A Corpus-Based Critical Discourse Analysis of Gender Representation in ELT Textbooks

Adela Rifa Anbiya¹, Fadillah Meigiansyah Rizqie¹, Luthfi Luqmansyah¹, Nazmara Azka Hidayat*¹, Hapid Ali¹

¹Universitas Islam Negeri Sunan Gunung Djati Bandung, West Java, Indonesia

Abstract

The subject of this study is the representation of gender in government compulsory English textbooks for Grade 4 students in elementary schools in Indonesia, 'My Next Words Grade 4'. The research on gender bias in EFL materials has been conducted before, but there is a lack of quantitative critical discourse research on the new textbooks that were introduced at the primary education level, namely the Kurikulum Merdeka textbooks. For this research, a corpus-assisted critical discourse analysis (CDA) technique is used to identify and quantify the gender-related references in the texts, such as the pronouns, the names of characters, and the visual cues of social and domestic roles. The data were quantified using the software AntConc to gain insight into the occurrence of male and female references and the collocations that arose. The results reveal unequal visibility and roles of genders in characters, with stereotypical representations remaining even though the government has been promoting gender parity in its media content, especially the association of domestic activities with female characters and active, public-sphere roles with male characters. The results of this study bring to the fore the ideology embedded in official learning materials and indicate that curriculum developers need to make more balanced and nondiscriminatory representations to create equitable gender perceptions among young learners.

Keywords: Critical Discourse Analysis, Gender Representation, Corpus-assisted, ELT Textbooks, Primary Education

1 Introduction

This study examines the gender disparity in the Grade 4 Merdeka Curriculum textbook, My Next Words. Despite considerable studies on gender representation (Adawiyah & Oktavianti, 2023; El Moudden & Said, 2025), quantitative research employing corpus methodologies is still scarce in Indonesian government primary school textbooks. This work uses Corpus-Assisted Critical Discourse Analysis (CDA) to address this gap. The use of corpus methods offers enhanced rigor compared to conventional qualitative approaches, which are occasionally critiqued for their methodological deficiencies (Cheng, 2012). Prior research indicates that textbooks frequently embody conventional stereotypes, both visually and vocally (Fathiyousefabad & Yangın Ekşi,

*Corresponding Author: nazhprod@gmail.com

2025; Herdiawan et al., 2025). This research's originality resides in the analysis of data from the recently adopted curriculum to assess the actual attainment of inclusive narratives.

2 Methods

This study employs a qualitative research design, specifically utilizing Corpus-Assisted Critical Discourse Analysis (CDA) to investigate gender representation in elementary school English textbooks. The theoretical framework is developed mainly from the three-dimensional model of CDA, developed by Fairclough (2015), in order to analyze the construction of power relations and gender ideologies within the discourse. Fairclough's (2015) three-dimensional framework encompasses: (a) description, which analyzes the linguistic characteristics and vocabulary employed in the text; (b) interpretation, which investigates the relationship between the text and its production and consumption processes; and (c) explanation, which contextualizes the discourse within wider socio-cultural and ideological frameworks. This methodology is especially pertinent for examining gender representation, as it enables researchers to transcend superficial linguistic patterns and reveal the foundational ideas that influence and are influenced by textual discourse.

This theoretical framework is enhanced by Baker's (2006) corpus linguistics technique, which provides a macro-level quantitative foundation to ensure the rigor of empirical research. The integration of corpus linguistics with CDA, known as Corpus-Assisted Critical Discourse Analysis, operates through the following scope and workflow: Initially, corpus software (AntConc) is employed to discern objective language patterns from the textual data, encompassing the frequency of gender-related references and the thematic collocations of these references (e.g., pronouns, nouns). Secondly, these quantitative findings provide empirical evidence that can be systematically analyzed to reveal patterns of representation that might not be immediately visible through qualitative analysis alone. Third, the textual evidence is then analyzed qualitatively using Fairclough's (2015) CDA framework in order to interpret the social and ideological implications of these patterns. Fourth, regarding multimodal analysis, the visual materials (pictures) are examined through the visual grammar framework established by Kress and van Leeuwen (2006), facilitating a systematic evaluation of the visual positioning of characters, their activities, spatial dominance, and social roles. This extensive methodology provides a detailed examination of the co-construction of gender representation in verbal and visual texts. This study seeks to combine quantitative corpus methods with qualitative interpretive analysis to achieve empirical rigor and profound ideological understanding of the discursive creation of gender.

2.1 Data Source

The primary data source for this study is the official English textbook for Grade 4, entitled *My Next Words Grade 4* published in 2021. This textbook is a government-mandated educational resource implemented under the Kurikulum Merdeka in Indonesia, serving as the primary material for English language instruction at the primary level. This textbook is one of the government-mandated educational materials that are used in the implementation of Kurikulum

Merdeka in Indonesia, for English language learning in primary school as the main educational material.

The secondary sources include pertinent peer-reviewed academic journals, scholarly articles, and theoretical literature. These secondary sources are particularly related to Corpus-Assisted Critical Discourse Analysis (CDA), representation of gender in EFL learning resources, and corpus linguistics. These journals are necessary not just to build up a comprehensive theory, but also to offer a comparative analytical base to critically work out the ideological implications gleaned from the primary data.

2.2 Data Collection and Analysis

Data collection was conducted by converting the PDF version of the textbook into a machine-readable text format. Subsequently, AntConc version 4.4.1, a corpus linguistics tool, was employed to analyze the corpus of My Next Words Grade 4. The analysis followed three systematic steps:

1. **Quantitative Extraction:** The Concordance tool of AntConc was employed to ascertain the frequency of gender-related allusions within the textual corpus of the "My Next Words Grade 4" textbook. This involved the extraction of pronouns (e.g., "he," "she") and familial nouns (e.g., "father," "mother") to ascertain the overall representation of male and female characters in each chapter. The case sensitivity parameter was disabled to maximize information retrieval from the dialogues and reading sections in the book.
2. **Collocation Analysis:** The Collocate tool was used to identify the verbs and adjectives that were statistically found around the gender words that refer to the target word in the textbook. The purpose of this process is to expose the different roles, tasks, and daily life activities of male and female characters in the book, as they are specifically designated in the narrative of the book in their home sphere or public sphere.
3. **Critical Interpretation:** The data obtained from the textbook were analyzed based on Fairclough's (2015) three-dimensional approach which consists of: (a) Description (analyzing the specific linguistic features and vocabulary used in a textbook, "My Next Words Grade 4"), (b) Interpretation (connecting the production of a text with the pedagogical discursive practice in elementary education, namely the teaching of English), and (c) Explanation (placing the results of the analysis in the broader socio-cultural context, including the evaluation of the textbook "My Next Words Grade 4" with respect to its inclusive objectives as laid out in the Kurikulum Merdeka).

3 Results

3.1 The Quantitative Landscape of Gender Representation

The corpus analysis of My Next Words Grade 4 reveals a clear quantitative imbalance in gender visibility. As shown in Table 1, male-related references ("he") appear with a frequency of 23, significantly outnumbering female-related references ("she"), which appear only 5 times.

Table 1: Frequency of Gender-Related References in My Next Words Grade 4

No.	Gender Reference	Frequency	Percentage (%)	Discursive Representation
1	He	23	53.5	Represents active and visible male characters in public contexts.
2	She	5	11.6	Represents female characters with relatively limited visibility.
3	Mother	10	23.3	Frequently associated with caregiving and domestic responsibilities.
4	Father	5	11.6	Commonly represented as a provider or authority figure.
	Total	43	100	

Note. Frequencies were generated using AntConc (Version 4.4.1) from the textual corpus of My Next Words Grade 4.

Furthermore, the role-based references indicate a distinct pattern: "mother" is mentioned 10 times, whereas "father" appears only 5 times.

These findings suggest that while the textbook attempts to include female characters, male subjects remain the primary agents in the overall discourse, maintaining a higher visibility in terms of textual pronouns. This answers the first research question regarding the quantitative visibility of gender in the textbook, which reveals that male characters are represented nearly five times more frequently than female characters in the textual corpus.

3.2 Discursive Roles and Ideological Construction

The collocation analysis provides deeper insights into the roles associated with each gender in My Next Words Grade 4. The pronoun "he" is statistically linked to public and active collocates such as at, gets, always, bike, goes, school, and football. This linguistic clustering frames the male character as a disciplined, mobile, and public-oriented subject. According to Fairclough's (2015) ideological analysis, these lexical selections embody a patriarchal ideology that defines masculinity through movement, productivity, and public involvement. The term "goes" signifies movement and agency, whereas "school" and "football" situate the male subject within institutional and recreational public domains. The recurrent employment of "always" and "gets" establishes a narrative of habitual discipline and regularity, promoting the ideological belief that male characters are inherently engaged and productive societal members.

Conversely, the pronoun "she" is strongly associated with cooking in the My Next Words Grade 4. This discovery corresponds with Adawiyah and Oktavianti (2023), who observed that conventional household roles for female characters endure even in contemporary curriculum resources. Although Maesaroh et al. (2022) identified a growing effort in Indonesian primary school textbooks to represent professional work roles equitably, our study indicates that the linguistic representation of "she" is still limited to domestic or stereotyped contexts. Through Fairclough's (2015) ideological lens, the association of "she" exclusively with "cooking" represents a discursive strategy that perpetuates traditional gender ideologies by naturalizing the connection between femininity and domestic labor.

The collocation analysis reveals that the pronoun 'he' is statistically associated with high-

Table 2: Collocational Patterns of the Pronoun "He"

Collocate	Rank	FreqLR	FreqL	FreqR	Range	Likelihood	Effect
at	1	15	6	9	1	33.953	2.823
gets	2	7	1	6	1	33.367	4.746
always	3	7	2	5	1	31.709	4.576
bike	4	7	1	6	1	26.639	4.046
goes	5	8	2	6	1	23.212	3.354
by	6	8	2	6	1	22.611	3.295
does	7	7	4	3	1	21.584	3.498
to	8	13	7	6	1	21.337	2.300
eating	9	4	1	3	1	20.102	4.939
school	10	10	4	6	1	18.082	2.454
cooking	11	3	3	0	1	17.596	5.524
up	12	7	1	6	1	17.567	3.046
football	13	2	1	1	1	12.965	5.939

Table 3: Collocational Patterns of the Pronoun "She"

Collocate	Rank	FreqLR	FreqL	FreqR	Range	Likelihood	Effect
cooking	1	4	1	3	1	38.350	8.141
is	2	8	3	5	1	21.100	3.074
playing	3	3	3	0	1	21.062	6.404
she	5	2	1	1	1	12.360	5.819

mobility and public-sphere activities, evidenced by collocates such as 'bike', 'goes', 'school', and 'football'. Furthermore, its frequent association with temporal and disciplinary markers like 'at', 'gets', 'always', and 'does' constructs a discourse where the male subject is framed through productive, public-oriented routines.

In sharp contrast, the pronoun 'she' displays a narrower array of collocates, with 'cooking' serving as the predominant association term. This language clustering indicates that, despite the textbook's aim to promote equitable engagement, female subjects are mostly shown through domestic chores, so perpetuating a conventional gendered division of labor within the textbook's narrative.

This clear distinction between the public-oriented 'he' and the domestic-oriented 'she' serves as a discursive strategy that maintains stereotypical gender identities, even within the contemporary framework of the Kurikulum Merdeka. This answers the second research question regarding how gender is discursively constructed through lexical choices and collocational patterns in the textbook.

3.3 Gender Bias in Dialogues and Reading Passages

Beyond the pronominal and collocational analysis, further examination of the My Next Words Grade 4 textbook reveals additional instances of gender bias embedded within dialogues and reading passages. For instance, in numerous interactions, male characters are consistently depicted as the initiators of conversation, offering directions, issuing instructions, and assuming leadership roles in group activities. The male characters are commonly shown as inquirers,

problem-solvers, and facilitators of dialogue, whereas female characters are more frequently depicted as passive responders, acquiescent, or emotive. This tendency promotes the ideological portrayal of males as dominant and females as subordinate, consistent with Fairclough's (2015) notion of ideological discourse that normalizes power dynamics through quotidian language usage.

The reading passages in *My Next Words Grade 4* frequently portray male characters partaking in outdoor activities like playing football, riding bicycles, and attending school independently, while female characters are predominantly depicted in indoor environments, performing household tasks or engaging in nurturing and caregiving activities. This spatial division of activities reflects a broader ideological discourse that associates masculinity with the public sphere and femininity with the private sphere. According to Fairclough's (2015) three-dimensional model, this exemplifies the ideological aspect of discourse, wherein textual selections perpetuate and validate prevailing social structures and power imbalances.

Moreover, the employment of occupational terminology in the textbook indicates that male characters are predominantly linked to professional and authoritative responsibilities, whereas female characters are typically shown in caregiving, educational, or domestic capacities. This occupational segregation serves to normalize gender-based divisions of labor and may implicitly influence young learners' perceptions of appropriate career paths based on gender. This finding is particularly significant given that the textbook is used at the primary education level, where children's gender identities and aspirations are still being formed.

4 Discussion

This study's findings indicate significant ideological implications concerning gender representation in the *My Next Words Grade 4* textbook. The numeric disparity between male and female references suggests that, despite the government's professed dedication to gender equality in education, the textbook favors male representation over female representation. This illustrates Fairclough's (2015) concept of the ideological function of discourse, wherein language selections are not impartial but instead perpetuate and reinforce prevailing power dynamics. The textbook's emphasis on male characters as predominant subjects of discourse unintentionally conveys that male experiences and viewpoints are prioritized over those of females.

Secondly, the collocational patterns of "he" and "she" demonstrate a systematic gendering of social positions and activities. The linkage of "he" with public activities like attending school, playing football, and cycling establishes a narrative of masculinity characterized by activity, mobility, and involvement in the public realm. In contrast, the association of "she" with domestic activities such as cooking constructs a discourse of femininity as passive, confined to the private sphere, and oriented toward domestic caregiving. This discursive construction is particularly problematic because it naturalizes traditional gender roles, presenting them as normal and inevitable rather than as socially constructed and changeable.

This study is distinguished from prior research by its comprehensive lexical analysis aimed at identifying specific instances of gender bias in *My Next Words Grade 4*. This study offers a detailed examination of collocational patterns and lexical associations that elucidate the embedding of gender ideologies within Indonesian textbooks, building upon the broader patterns

identified in earlier studies by Adawiyah and Oktavianti (2023) and Maesaroh et al. (2022). This study rigorously applies Fairclough's (2015) three-dimensional framework to each lexical item, illustrating the connection between superficial linguistic choices and the underlying ideological structures that influence students' comprehension of gender.

Additionally, this study reveals that the gender bias in *My Next Words* Grade 4 is not limited to pronominal references but extends to the distribution of activities, roles, and spatial positioning of characters within dialogues and reading passages. This thorough analysis, integrating quantitative corpus approaches with qualitative critical discourse analysis, offers a more substantial critique of the textbook's gender depiction than earlier studies that depended exclusively on qualitative content analysis. The findings indicate a continual disparity between the curriculum's declared inclusive goals and the actual textual practices in the textbook, implying that policy commitments alone are inadequate for attaining gender equality without meticulous consideration of the linguistic and discursive specifics of educational materials.

The theoretical ramifications of these findings are substantial for comprehending the influence of educational discourse on the formation of gender ideas. Utilizing Fairclough's (2015) notion of ideological discourse, the textbook serves as a locus for the production, reproduction, and legitimization of prevailing gender beliefs via quotidian language usage. By consistently associating male characters with public-sphere activities and female characters with domestic roles, the textbook participates in what Fairclough calls the "naturalization" of ideology, where socially constructed power relations are presented as common sense or natural.

El Moudden and Said (2025) observe that CDA research regularly indicates that textbook speech can implicitly influence students' perceptions of gender-appropriate behavior. The textbook's portrayal of female pupils engaged in household duties may unintentionally support conventional Indonesian cultural standards, which, as noted by Ahmadabadi and Azad (2023), can limit the self-efficacy of young female learners concerning their future career trajectories. This is especially troubling considering that the pupils using *My Next Words* Grade 4 are at a critical developmental stage in their comprehension of gender roles and their own potential.

Furthermore, the results of this study contest the presumption that contemporary curriculum materials inherently result in more fair portrayals. Notwithstanding the progressive objectives of the Kurikulum Merdeka, the examination of *My Next Words* Grade 4 indicates that archaic patriarchal attitudes endure in nuanced yet substantial ways. This indicates that curriculum makers and textbook authors must conduct more meticulous and systematic gender audits of educational materials to ensure alignment with the values of gender equality and inclusivity advocated by the curriculum.

5 Conclusion and Recommendation

5.1 Conclusion

This study shows that, despite the introduction of the Kurikulum Merdeka, intended to foster inclusive and contemporary education, gender representation in the *My Next Words* Grade 4 textbook continues to be shaped by conventional patriarchal views. The quantitative analysis indicates a substantial visibility disparity, with male characters ("he") depicted as primary agents

about five times more often than female characters ("she"). This finding clearly addresses the initial research question concerning the numerical representation of gender in the textbook, affirming that male characters predominate in the textual discourse.

Moreover, the collocation analysis offers discursive evidence that male characters are linked to public-sphere mobility and structured activities such as attending school, playing football, and cycling, whereas female characters are consistently relegated to domestic roles, notably "cooking. The additional analysis of dialogues and reading passages reveals that male characters are more frequently portrayed as authoritative, active, and engaged in outdoor or professional activities, while female characters are depicted as passive, nurturing, and confined to domestic spaces. These findings address the second research question regarding the discursive construction of gender through lexical choices and contextual positioning.

Consequently, the textbook discourse inadequately attains gender equity, as it tacitly reinforces stereotypical labor divisions that reflect conventional Indonesian society norms. The ideological analysis employing Fairclough's (2015) paradigm demonstrates that these linguistic patterns function to normalize and validate patriarchal power structures, potentially shaping young learners' views on gender-appropriate conduct and future ambitions.

5.2 Recommendation

Future curriculum developers are advised to perform a comprehensive gender audit on textbook manuscripts, guaranteeing equitable visibility and diverse, non-stereotypical roles for both female and male characters. English educators must adopt a critical pedagogical approach while using these materials, aggressively confronting gender-biased content during classroom discussions. Future researchers are urged to enhance this study by integrating a multimodal approach, investigating the interaction between visual illustrations in the textbook and textual allusions to either reinforce or undermine these gendered biases.

References

- Adawiyah, S. Z., & Oktavianti, I. N. (2023). Gender representation in merdeka curriculum elt textbooks: A corpus-assisted study. *Indonesian Journal of EFL and Linguistics*, 8(2), 313–326. <https://doi.org/10.21462/ijefl.v8i2.668>
- Ahmadabadi, Z. S. H., & Azad, M. (2023). The critical discourse analysis of gender representation in english textbooks of iranian high schools: The fairclough model in focus. *Conhecimento & Diversidade*, 15(38), 413–442.
- Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.
- Cheng, W. (2012). Corpus-based linguistic approaches to critical discourse analysis. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (1st ed.). Wiley. <https://doi.org/10.1002/9781405198431.wbeal0262>
- El Moudden, R., & Said, K. (2025). Locating critical discourse studies within school textbook research: A systematic literature review of research articles. *Cogent Education*, 12(1), 2481006. <https://doi.org/10.1080/2331186X.2025.2481006>

- Fairclough, N. (2015). *Critical discourse analysis: The critical study of language* (2nd ed.). Routledge.
- Fathiyousefabad, M., & Yangin Ekşi, G. (2025). Gender representation and bias in visuals in local elt course books: A critical discourse analysis. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 45(2), 587–604. <https://doi.org/10.17152/gefad.1704587>
- Herdiawan, R. D., Fakhruddin, A., Nurhidayat, E., & Nurhadi, K. (2025). Gender stereotypes in visual and verbal texts from government-distributed efl textbooks: Critical discourse analysis. *Journal on English as a Foreign Language*, 15(1), 367–393. <https://doi.org/10.23971/jefl.v15i1.9357>
- Kress, G., & Van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203619728>
- Maesaroh, D. T., Asih, Y. U., & Suhatmady, B. (2022). Does the english textbook for young learners in indonesia promote gender equality? *IJELTAL (Indonesian Journal of English Language Teaching and Applied Linguistics)*, 7(1), 35. <https://doi.org/10.21093/ijeltal.v7i1.1152>