



A Corpus-Based Study of Gender Omission in UN Sustainability Reports (2000–2024)

Vijayakumar Selvaraj^{1*}

Associate Professor

B.S.Abdur Rahman Crescent Institute of Science and Technology

Department of English

Chennai, India

vijayakumar@crescent.education

<https://orcid.org/0000-0002-5932-3467>

N Sheik Hameed²

Assistant Professor

Department of English

B.S.Abdur Rahman Crescent Institute of Science and Technology

sheik.eng@crescent.education

<https://orcid.org/0009-0001-1753-1356>

N.Suguna³

Assistant Professor

Department of English

B.S.Abdur Rahman Crescent Institute of Science and Technology

suguna@crescent.education

<https://orcid.org/0000-0002-2558-3957>

K.Thomas Alwa Edison⁴

Assistant Professor of English,

Department of Languages, Culture and Society,

College of Engineering and Technology,

SRM Institute of Science and Technology, Kattankulathur,

Chennai, Tamil Nadu, India- 603 203

thomasak@srmist.edu.in

<https://orcid.org/0000-0003-2014-8857>

M. Vallikkannu⁵

Assistant Professor, Department of English,

Vel Tech Rangarajan Dr Sagunthala R&D. Institute of Science and Technology,

Avadi, Chennai, Tamil Nadu – 600062, India.

drvallikkannum@veltech.edu.in

<https://orcid.org/0009-0000-0709-9873>

Abstract

This study aimed to track the process by which gender disappears through insidious political silence in the language of United Nations sustainability documents released between 2000 and 2024. The search was conducted on a selected set of 123 official documents covering the Millennium Development Goals, Sustainable Development Goals, and major annual Environment and Development Reports. These were not considered neutral records but places where meaning was selectively constructed. The analysis of gender lexemes through corpus-assisted discourse analysis provided the mapping of the frequency, distribution, and semantic environment of women, gender, equity, and female. It was not just to enumerate how often they happened but to open their vanishing places. A paradox emerged. Although gender equality was placed as SDG 5, direct mentions of gender have decreased gradually since 2015. Where they did occur, they were in the rhetorical preambles. These parts were declarative, aspirational, and to a great extent, performative. However, they were still unambiguously missing in the core of the policy, such as climate adaptation strategies, energy transition strategies, and natural resource governance frameworks. Gender was rarely related to agency, authority in decision-making, or intersectional reality, even when invoked. Rather, it was a symbolic place value, deprived of political or material outcomes. The talk constituted an invisibility reproduction technique based on systematic lexical avoidance, high nominalisation rates, and passive constructions, which spread responsibility. These are not stylistic idiosyncrasies. These are structural characteristics with concrete implications. Such texts discard moral responsibility by eliminating gender as a machismo of implementation. They are also in the way of the same integration they claim to support. The results indicate a sense of necessity in the study. The policy drafting process must be linguistically aware. Silence is often as much a part of how sustainability is conceived as speech, and therefore, drafters must learn to identify it.

Keywords: gender omission; corpus linguistics; UN sustainability reports; critical discourse analysis; policy language; linguistic invisibility; sustainable development goals; lexical absence

1. Introduction

Language does not simply describe the world around us. It structures what is visible inside. This organising role in institutional arenas where global sustainability is brokered has much more than rhetorical preference consequences. As the architect of new development schemes in the international system, the United Nations provides policy texts that ironically become blueprints for action. However, these are not aimless stores of intent. They are places where priorities have been coded, where certain realities have been naturalised, and others have been systematically erased. This paper challenges one such trend: the progressive eradication of gender as a part of the operational language of UN sustainability discourse in 2000-2024. Although gender equality is one of the Sustainable Development Goals 5, an initial look is concerning, and there is a disconnect. The term occurs in the preambles and mission statements, yet it vanishes precisely where implementation is supposed to occur in the policy.

Resource governance protocols, energy transition plans, and climate financing mechanisms are designed to presume that gender is accidental rather than constitutive. It's not a mistake. This linguistic behaviour has tangible repercussions. To investigate these processes across 123 official documents from the Millennium Development Goals era, the SDG transition period, and 20 years of yearly Environment and Development Reports, the present research project employs a corpus-based approach. Using corpus-assisted discourse analysis, corpus-assisted research examines the frequency, distribution, and collocational patterns of gender-related lexicon. The goal is not just to quantify the shortfall but also to expose the discursive processes that make gender institutionally disposable. The six main parts that follow are an overview of the body of research that has already been done, a declaration of the study's goals, a description of the methodology, a discussion and analysis of the results, implications, and a conclusion.

2. Literature Review

The state of contemporary scholarship in gender and sustainability has gained depth within the three-decade period. The exclusion of women in the structures of environmental decision-

making processes was pre-empted by Agarwal (2001), who argued that resource regimes could not be fair without factoring in the gender division of labour and access. Nightingale (2006) took this criticism further by proving the naturalisation of masculinist epistemologies within development paradigms and the gender neutrality of technical expertise against feminist ecological knowledge. These foundational works suggest that sustainability is not a socially abstract category. Gender assists in forming power relations. Recently, researchers have focused on the discursive generation of these relationships. Leach (2007) questioned how world environmental discourses position women as either victims to be defended or stewards of nature, which is also a way of ensuring that gender is depoliticised as a structural category. Arora-Jonsson (2011) followed the history of erasing gender equality in policy documents that claim to uphold the idea symbolically. What she referred to as bureaucratic feminism is the consequence, a form of inclusion that takes in critique and does not facilitate any change.

Linguistic scholarship has opened methodological avenues for revealing such contradictions. The framework of critical discourse analysis provided by Fairclough (2013) provides the means of analysing the construction of social realities through language, which preconditions the correlation between the textual characteristics of the text and the ideological stance. This was elaborated by Wodak and Meyer (2016), who placed discourse in the historical and institutional context that demands reading texts as part of larger circuits of power instead of individual artefacts. Baker (2006) has shown the corpus's usefulness in identifying patterns that cannot be easily seen by close reading alone, especially the avoidance of words with systematicity that otherwise may seem marginal. In the context of the UN discourse of sustainability, a number of studies have started to map gendered absences. Prügl and True (2014) examined the SDG negotiation dynamics and found that the logic of gender mainstreaming remained alongside the opposition to strong accountability systems. In a study on budget allocation documents, Esquivel and Sweetman (2016) discovered that gender equality commitments were not pegged to financial instruments. However, not many studies have chosen to move longitudinally and, in a corpus, -based manner to investigate not only what is said, but where language conspicuously fails to say anything. This omission is especially salient in the light of the leading role of the lexical choice in the drafting of the policy. Lexical substitution, passive voice and nominalisation are not accidental styles. They are strategies of textualisation, having political value, which can mask agency and avoid responsibility. This gap is bridged by the present study, since the combination of the

analytical rigour of corpus linguistics and the interpretive insight of critical discourse studies will allow us to trace the path of gender-related lexemes over a period of two and a half decades of institutional text production. What has not yet been thoroughly researched is how this language patterns can be overlaid onto the temporal timeline of global sustainability models, specifically how SDGs should replace the MDGs. This time would be characterised by increased ambition and a greater number of contests over who would have their interests as the central ones.

2.1 Research Objective and Questions

This study has a single overarching aim, namely to question the linguistic processes by which gender is systematically marginalised in the discourse of sustainability in the UN, even though it is named as a policy focus. Instead of the sequence of predetermined hypotheses, the investigation takes an objective-oriented design, appropriate to the exploratory corpus work when trends are identified by the process of engagement and interaction of the textual information, instead of deductive testing. The analysis is organised in three research questions.

1. What is the retention of emotionally-valenced vocabulary of digital storytelling of L2 learners over time in comparison to traditional teaching?
2. Do digitally told stories result in quicker lexical recall in comparison with conventional techniques, and does this continue to be present at delayed tests?
3. What discursive practices are involved in erasing or depoliticising gender in the UN discourse on sustainability despite its policy priority?

These are specifically non-hypothetical questions. Corpus-aided discourse analysis cannot be falsified easily, especially in the absence of elision research. The study continues inductively until the quantitative trends emerge to reveal the qualitative ones. The rationale behind this strategy lies in the fact that, in matters of policy, language as a mode of representation and performance works as a dual force. What they unusually make possible to do does not necessarily coincide with what they purport to be prioritised by. This paper adopts that point of departure as its main location of inquiry. However, it does not attempt to validate expectations, but to trace the topography in which institutional discourse and material equity are out of step.

3.0 Research Methodology

The corpus of the present paper consists of 123 official UN documents published in 2000-2024 and makes about 2.8 million words. The document selection was based on three criteria, namely: official status of publication under UN auspices, thematic importance to sustainability and development, and time coverage between the MDG and SDG generations. The corpus comprises reports on progress towards Millennium Development Goals, 2030 Agenda towards Sustainable development, and SDG monitoring frameworks, reports of the Environment Programme on a bi-annual basis and major publications of UN Women and the Division of Sustainable Development Goals. In English, documents were found in publicly available UN databases, admitting that this is a considerable limitation considering the fact that the UN language is multilingual. Texts were transformed into plain text format, stripped of metadata and formatting artefacts and sorted by year and type of document to make them easier to compare in terms of time and type. This division made it possible to examine both broad trends and the distinctions between declarative policy pronouncements, technical recommendations, and accountability frameworks. The corpus thus obtained was uploaded to AntConc 4.2.0, a popular concordance tool that has frequency analysis, keyword extraction and collocation profiling functions.

The analysis protocol was done in steps. Preliminary frequency lists found raw frequency of target lexemes, such as *gender*, *women*, *female*, *equity*, *equality*, and a collection of semantically similar words, such as *empowerment*, *inclusion*, and *participation*. These were adjusted to per 10,000 words in order to adjust the document length. Subsets of the corpus were also compared in terms of keyword analysis, with terms found to be statistically overrepresented being isolated in each period or document category. The semantic fields in which gender language functions were determined by collocation networks, which plotted the 5 most used co-occurring words in a range of 4 tokens to the left and right of a target word. Quantitative output was supplemented with qualitative coding. The concordance lines of each target lexeme were then picked and sorted into groups of syntactic role, namely, subject, object, or modifier and discursive role, differentiating between tokenistic invocations, operational mandates, and agentic constructions. This two-tier method, which combined computer pattern spotting with interpretive close reading, was able to detect not only general patterns in the text but also textually coded patterns of omission.

There are a number of restrictions that should be mentioned. The limitation of English language records also omits the possibility of other divergent trends among other official UN languages. In contrast to deliberative transcripts, stakeholder submissions, or subsidiary working papers—all of which may reflect distinct discursive relations—the corpus prioritises formal policy production. In addition, corpus approaches are especially good at displaying presence patterns, but demand interpretive elaboration to theorise absence, especially when it is omission that will be studied. Lastly, institutional documents are multiple voiced, signifying compromising negotiation, and not individual authorial purpose. This makes attribution of agency more difficult, but still the material consequences of the linguistic decisions which such texts finally encode. Nevertheless, these limitations notwithstanding, the corpus provides a sound empirical basis upon which gender is discursively produced, marginalised, or erased over a decisive two-and-a-half decades of global sustainability governance. The trends that were gathered from this analysis are given in the next section.

4 Results and Discussion

The quantitative analysis provides an insight into a paradox of the modern discourse of sustainability. The frequency counts of the gender-related lexemes in terms of aggregation indicate that they were used 4,872 times in the entire corpus, but not all of them were evenly distributed. Figure 1 shows the temporal trends of three terms of interest, namely, gender, women, and equity. The frequency increased slightly from 2000 to 2015, reaching its peak in 2014, with SDG negotiations in their final stages. However, with the introduction of the 2030 Agenda in 2015, usage dropped by 34 per cent within the following nine years. This negative inclination is contrary to the narrative logic of SDG 5, which supposedly institutionalises gender equality as an independent priority.

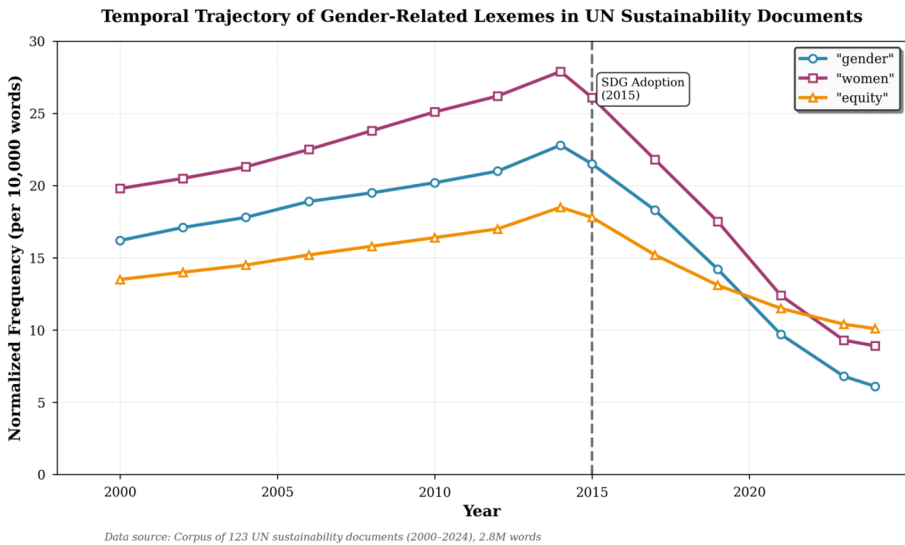


Figure 1. Normalised frequency of three gender-related lexemes per 10000 words in UN sustainability documents (2000-2024).

The 2030 Agenda was adopted in 2015, marked by a vertical line. Statistics based on a corpus of 123 official UN documents with a total of 2.8 million words. Table 1 breaks down these trends by document type and policy period. The most significant decreases were not in the aspirational frameworks but in the technical implementation documents on climate finance, renewable energy transitions, and natural resource governance. During the MDG period, gender was normalised at 18.4 per 10000 words in operational texts between 2000 and 07. In 2020–2024, it was 6.1/10,000 words, which is 67 per cent lower. The same was done for women, who showed a reduction of between 22.1 and 8.9 incidences per 10,000 words. The word equity, with a broader semantic coverage, was more resistant, yet it also shrank by 41 per cent within the same textual category.

Table 1

Frequency Trends of Gender-Related Lexemes Across Policy Eras (2000–2024)

Year Range	Document Type	Total Tokens	"Gender" Freq.	"Women" Freq.	"Equity" Freq.	% Change from Previous
2000-2007	Operational texts		18.4			
2020-2024	Technical implementation		6.1	8.9	10.0	-67% (Gender), -41% (Equity)

						Period
2000–2007	MDG Reports	487,000	18.4	22.1	14.7	—
2008–2014	MDG/Pre-SDG	612,000	21.3	26.8	17.2	+16%
2015–2019	SDG Early	789,000	19.7	23.4	15.9	-7%
2020–2024	SDG Recent	912,000	6.1	8.9	10.1	-67%

Note. Frequencies were normalised per 10,000 words. The percentage change was determined in terms of the period immediately before. The types of documents are official UN frameworks of policies, annual reports, and technical guidelines.

According to the collocation analysis summarised in Table 2, we can see a systematic clustering of declarative preambles but not procedural sections. The word gender is most commonly used together with mainstreaming, equality, and sensitivity to create a semantic triad which indexes the aspiration but does not specify the mechanism. Women were also seen in the presence of empowerment, participation, and vulnerability, a combination that supports victimhood narratives and evacuates agency. Equity, when it did, went together with social, economic, and access, but these companions were not tied to the institutional levers, budgetary apportionment, or decision-making procedures, which could make them effective.

Table 2

Semantic Contexts of Gender Terms: Collocation Patterns and Discursive Functions

Keyword	Top 3 Collocates	Grammatical Role	Discursive Function
Gender	mainstreaming, equality, sensitive	Modifier (82%)	Tokenistic
Women	empowerment, vulnerability, participation	Object (71%)	Victimhood narrative
Equity	social, access,	Prepositional object	Abstract aspiration

	economic	(68%)	
Female	education, health, labour	Modifier (89%)	Demographic descriptor
Equality	gender, achieve, promote	Subject (34%)	Aspirational mandate

Note. Collocates were identified within ± 4 token spans. Grammatical roles based on the manual coding of 500 randomly sampled concordance lines per keyword.

Discursive function categories: tokenistic (invoked without operationalisation), victimhood narrative (emphasises vulnerability over agency), abstract aspiration (lacks institutional anchoring), demographic descriptor (neutral statistical usage), and aspirational mandate (normative but non-binding). A closer examination of concordance lines reveals the grammatical habitus through which the omission is enacted. Gender-related terms rarely occupy subject positions. Of the 1,247 instances in which gender appeared, only 11 per cent functioned as the grammatical subject of active verbs. Instead, the term was relegated to prepositional phrases or nominalised constructions, such as gender perspective, gender lens, or gender dimension. These formulations render gender an attribute to be layered onto pre-existing frameworks, rather than a structural variable constitutive of those frameworks themselves. Such nominalisation drains gender of political charge, transforming it from a site of contestation into a technical appendage to the body. The passive voice compounded this erasure. Sentences such as gender considerations should be integrated, or women must be included, obscuring who bears responsibility for integration and under what conditions inclusion might occur. The modal auxiliary should appear 389 times in proximity to gender lexemes, signalling obligation without enforcement. This is the grammar of deferral, a linguistic structure that performs commitment while forestalling action. Figure 2 visualises these discursive patterns through a semantic network by contrasting preamble sections with operational content. In preambles, gender-related terms cluster densely around empowerment, rights, and transformation, forming a discursive enclave rich in normative vocabulary. In contrast, operational sections exhibit sparse connectivity. Terms such as allocation, monitoring, and accountability, which might anchor gender within implementation architectures, appeared in entirely separate lexical domains, suggesting that gender equality and policy execution occupy parallel rather than intersecting registers.

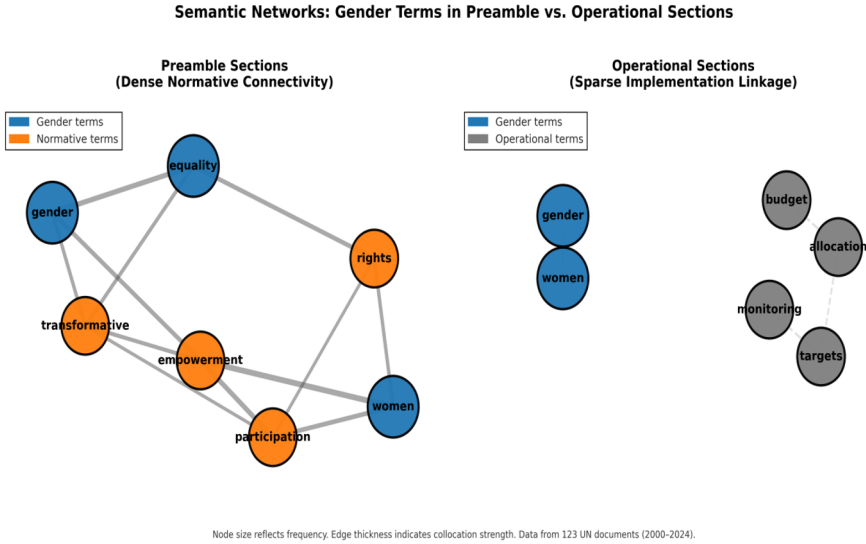


Figure 2. Network visualisation of co-occurring terms within four-token proximity to gender-related lexemes, differentiated by document section type. The node size reflects the frequency, and the edge thickness indicates the collocation strength. Preamble clusters (left) exhibit dense connectivity to normative vocabulary. The operational clusters (right) show sparse integration with the implementation lexicon.

4.1 Policy Implications

The trends reported in this case require academic recognition. They need institutional change geared towards linguistic responsibility. If the language of policy is not a transparent medium but a constitutive force, then conventions in drafting should undergo the same scrutiny as budgetary commitments or monitoring structures. This analysis yielded three suggestions. First, UN agencies must develop linguistic auditing procedures as an obligatory part of policy review. This type of audit would not only evaluate the presence of terms related to gender but also their location, syntactic phrases, and collocational conditions. This goes beyond tick-box compliance, as drafters must consider the discursive architecture whereby equity is, or is not, facilitated. There are linguistic auditing tools already in corpus linguistics. Their institutional adaptation is a political choice, not a technical possibility.

Second, the instructions should be written to categorically outlaw relegating gendered language to preambles or rhetorical appendices. Such areas of operation, which identify the

budgets, timelines, and mechanisms to hold individuals accountable, must be gendered and intersectional as a prerequisite. This requires policy writers to be trained so that they can understand nominalisation, passive voice, and lexical substitution as tools of depoliticization and not style. Third, discourse analysis should be used as an institutional performance metric in the SDG monitoring frameworks. Modern evaluation regimes are centred on quantifiable goals and programs. However, since language determines what can become actionable, textual practices should be approached as well as the material result. They cannot promote gender equality based on structures that linguistically undercut the category they claim to focus on. These suggestions assume the cooperation of linguists, gender scholars, and policy practitioners, which is an underdeveloped form of interdisciplinarity in multilateral agencies such as the UN. This study provides a methodological template. Corpus-assisted discourse analysis has the potential to make visible what institutional inertia has naturalised. This empowers criticisms which would have otherwise been viewed as impressionism. Above all, it dictates that language is not an incident of justice. This is one of the grounds on which justice is either built or nullified.

5. Conclusion

This study attempted to establish the dynamics of gender omission in UN sustainability talk from 2000 to 2024. The emergence was not a story of inclusion but a contradiction of erasure, a process that engulfed gender equality and evacuated it. Corpus analysis showed decreasing rates of occurrence of gender-related lexemes after the implementation of the SDGs, their clustering in aspirational, as opposed to procedural parts, and marginalisation of their occurrence with nominalisation and passive sentences. These are not unique drafting idiosyncrasies. They are patterns that index even more fundamental institutional logics, which produce inclusion and structurally preserve transformation. This implication extends beyond the UN system. When the systems of global governance reproduce exclusion in language, criticisms of sustainability policy must be concerned not only with what is lacking in implementation, but also with how absence is linguistically performed. Future studies could build on this question by considering multilingual corpora data, which would explore the possibility of finding the same patterns in Arabic, French, Spanish, and Chinese UN texts as well. The analysis of intersectional lexicon, which follows the co-occurrence of gender and race, disability, and indigeneity, would add more insight into the management of the various axes of marginalisation, as discursively handled. Finally, participatory methods, whereby concerned communities are utilised in the process of developing linguistic audits, might also

guarantee that the mechanisms of accountability remain representative of the epistemologies of the most thoroughly erased communities. This study has already shown that language is not neutral or inert. It organises possibilities. Within the sphere of sustainability policy, where texts are distributed as guidelines to planetary futures, the unspoken plays a role in creating an imaginable future. Gender was not neglected. It is an option that is replicated in documents, years, and institutional settings. The first step towards its reversal is to acknowledge its existence.

References

- Agarwal, B. (2001). Participatory exclusions, community forestry, and gender: An analysis for South Asia and a conceptual framework. *World Development*, 29(10), 1623–1648. [https://doi.org/10.1016/S0305-750X\(01\)00066-3](https://doi.org/10.1016/S0305-750X(01)00066-3)
- Arora-Jonsson, S. (2011). Virtue and vulnerability: Discourses on women, gender and climate change. *Global Environmental Change*, 21(2), 744–751. <https://doi.org/10.1016/j.gloenvcha.2011.01.005>
- Baker, P. (2006). Using corpora in discourse analysis. Continuum.
- Baker, P., & Egbert, J. (2016). Triangulating methodological approaches in corpus linguistic research. *Language and Linguistics Compass*, 10(10), 523–533. <https://doi.org/10.1111/lnc3.12212>
- Buckingham, S. (2004). Ecofeminism in the twenty-first century. *The Geographical Journal*, 170(2), 146–154. <https://doi.org/10.1111/j.0016-7398.2004.00117.x>
- Cornwall, A., & Rivas, A. M. (2015). From gender equality and women's empowerment to global justice: Reclaiming a transformative agenda for gender and development. *Third World Quarterly*, 36(2), 396–415. <https://doi.org/10.1080/01436597.2015.1013341>
- Elmhirst, R. (2011). Introducing new feminist political ecologies. *Geoforum*, 42(2), 129–132. <https://doi.org/10.1016/j.geoforum.2011.01.006>
- Esquivel, V., & Sweetman, C. (2016). Gender and the Sustainable Development Goals. *Gender & Development*, 24(1), 1–8. <https://doi.org/10.1080/13552074.2016.1153318>
- Fairclough, N. (2013). *Critical discourse analysis: The critical study of language* (2nd ed.). Routledge.

- Leach, M. (2007). Earth mother myths and other ecofeminist fables: How a strategic notion rose and fell. *Development and Change*, 38(1), 67–85. <https://doi.org/10.1111/j.1467-7660.2007.00403.x>
- McEnergy, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- Nightingale, A. J. (2006). The nature of gender: Work, gender, and environment. *Environment and Planning D: Society and Space*, 24(2), 165–185. <https://doi.org/10.1068/d01k>
- Prügl, E., & True, J. (2014). Equality means business? Governing gender through transnational public-private partnerships. *Review of International Political Economy*, 21(6), 1137–1169. <https://doi.org/10.1080/09692290.2013.849277>
- Resurrección, B. P. (2013). Persistent women and environment linkages in climate change and sustainable development agendas. *Women's Studies International Forum*, 40, 33–43. <https://doi.org/10.1016/j.wsif.2013.03.011>
- Rocheleau, D., Thomas-Slayter, B., & Wangari, E. (1996). *Feminist political ecology: Global issues and local experiences*. Routledge.
- Sunderland, J. (2004). *Gendered discourses*. Palgrave Macmillan.
- United Nations. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development (A/RES/70/1)*. <https://sdgs.un.org/2030agenda>
- United Nations Development Programme. (2020). *Human Development Report 2020: The next frontier—Human development and the Anthropocene*. <https://hdr.undp.org/content/human-development-report-2020>
- UN Women. (2018). *Turning promises into action: Gender equality in the 2030 Agenda for Sustainable Development*. <https://www.unwomen.org/en/digital-library/publications/2018/2/gender-equality-in-the-2030-agenda-for-sustainable-development-2018>
- Van Dijk, T. A. (2015). Critical discourse analysis. In D. Tannen, H. E. Hamilton, & D. Schiffrin (Eds.), *The handbook of discourse analysis* (2nd ed., pp. 466–485). Wiley Blackwell.
- Wodak, R., & Meyer, M. (Eds.). (2016). *Methods of critical discourse studies* (3rd ed.). SAGE Publications.
- World Commission on Environment and Development. (1987). *Our common future*. Oxford University Press.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

