

ANALYSIS OF ACADEMIC DISCOURSE ON SUSTAINABILITY THROUGH NATURAL LANGUAGE PROCESSING (NLP)

Nora Gavira-Durón¹, Luis Enrique García-Pérez¹, Muhammad Kashif¹,
Angelica Alonso-Rivera¹, & Ana Lorena Jiménez-Preciado²

¹Universidad de las Américas Puebla (Mexico)

²Instituto Politécnico Nacional CDMX (Mexico)

Abstract

This research examines the predominant conceptual frameworks in international academic literature on corporate sustainability and assesses their alignment with the principles of Environmental Education for Sustainability. Using a mixed-methods design applied to a corpus of 604 high-impact articles indexed in Scopus (2023-2024), we combined quantitative n-gram analysis through Natural Language Processing (NLP) with qualitative framing analysis. This approach enabled the identification of the most frequent terms and phrases (1, 2, and 3-word combinations), as well as the interpretation of their discursive context and underlying logic. The results reveal an academic discourse predominantly structured around corporate-financial logic. The marked prevalence of terms such as "sustainability," "company," and "corporate" among the most frequent 1-grams, together with the recurrence of bigrams like "financial performance" and "corporate social responsibility," demonstrates an approach that subordinates sustainability to management and profitability objectives. Contextual analysis confirms that these concepts are primarily framed as instruments for risk management and economic value maximization, marginalizing essential perspectives such as environmental ethics, climate justice, and planetary boundaries. It is important to acknowledge that the study is limited to English-language literature indexed in Scopus, which may exclude valuable perspectives published in other languages or regional journals. Furthermore, although n-gram analysis provides robust identification of discursive patterns, its quantitative nature requires complementary qualitative contextual analysis to adequately interpret underlying meanings. Based on these findings, the results highlight the need to incorporate "critical discourse literacy" into Environmental Education for Sustainability programs. We propose training future leaders in deconstructing the identified hegemonic narratives—which subordinate sustainability to financial logic—to foster more autonomous, critical, and systemically transformative ecological literacy.

Keywords: *Sustainability education, discourse analysis, Natural Language Processing (NLP), epistemic bias, finance and sustainability.*

1. Introduction

Environmental education is critical for the development of ecological literacy and promotion of responsible behavior. The primary focus of education for the environment is the development of environmental ethics and a sense of responsibility aimed at conservation. Environmental education provides the human and cultural values, skills, and system thinking needed to implement the corporate responsibility frameworks such as the Triple Bottom Line (TBL), Stakeholder Theory, and Resource-Based View (RBV). The connection between environmental education and corporate responsibility is often explored through organizational learning, corporate social responsibility (CSR) in education, and Education for Sustainable Development (ESD).

Academic discourse on sustainability combines environmental, social, and economic aspects within ESD. ESD focuses on long-term solutions for human well-being and interconnected challenges. Environmental Education for Sustainability (EEfS) provides an essential ecological foundation that serves as a pathway toward the broader framework of ESD. Academic discourse on sustainability has foundations in Sustainable Development Goals (SDGs), presented by the UN in 2015 and received global commitment by 193 countries on 17 goals with 169 targets for 2030. Within the SDG framework, Quality Education comes under pillar 4, that requires data, research, and educated professionals. EEfS uses modern education to implement the principles of the SDGs (UN, 2016). The World Economic Forum's 2019/2020 report also

emphasizes on sustainability through research and policy development, as the Global Competitive Index (GCI) includes higher education and training as Pillar 5 in its 12 pillars.

The Environmental Education for Sustainability (EEfS) is the transfer of knowledge, ethics, and behaviors essential for environmental sustainability. It is a value-driven education that requires lifelong learning, action, and reflection. EEfS is based on principles that represent the response of education to global ecological crises. The core principles of environmental education for sustainability can be categorized into four groups: Foundational, Pedagogical, Socio-Cultural, and Process-Oriented.

The foundational principles of EEfS have a philosophical basis in holistic and interdisciplinary, lifelong learning, and systems thinking approaches (Nasibulina, 2015; Tilbury, 1995). The holistic and interdisciplinary approach considers that education must integrate environmental perspectives across all subjects, so that the learners can detect connections and understand trade-offs between ecological and socio-economic systems (Tilbury, 1995). The lifelong learning process is the environmental education that includes formal settings and informal contexts like community programs and vocational training (Huckle & Sterling, 1996). Lastly, systems thinking is defined as the understanding of a series of complex and interconnected systems, that considers the environment in its totality of interdependent surroundings rather than just nature (Putilova et al., 2023).

The pedagogical approaches are used to apply foundational principles of EEfS. The pedagogical approaches include experiential and place-based learning, learner engagement and empowerment, and critical thinking and reflection. The experiential and place-based learning approach is direct, firsthand experience and learning in local landscapes, and community, using them as a foundation for study the curriculum. Learner engagement and empowerment occurs where learners actively engage in the teaching/learning process and improve decision making on sustainability issues. Lastly, the education should support critical thinking and reflection for learners to enable them to review cultural, economic, and political values and assumptions that may undermine the ecological systems (Tilbury, 1995). The socio-cultural dimensions and process-oriented principles are essential for inclusion and social justice, ensuring that education addresses environmental inequities and empowers all learners through participatory and collaborative approaches (UNESCO, 2020).

This study explores academic discourse on EEfS by characterizing key research and conceptual frameworks in international academic literature. The study uses qualitative framing analysis as a diagnostic tool to examine the academic discourse. It aims to identify the shortcomings in corporate sustainability literature using n-gram analysis through Natural Language Processing (NLP). It reveals the underlying values and assumptions and proposes critical discourse literacy as a pedagogical outcome of the critical thinking and reflection principle. Critical discourse literacy focuses on the educational outcome of the research. This discourse is necessary to equip individuals with the knowledge, skills, and values needed to address environmental challenges and promote sustainable development.

The rest of the paper is structured as following. Section 2 reviews the existing literature related to this article, while Sect. 3 presents the methodology. Section 4 focuses on results and discussion and finally, Sect. 5 concludes.

2. Literature review

The study of academic discourse on Corporate Social Responsibility (CSR) in sustainability is related to a number of other works in the literature, which can be grouped into three categories: (i) NLP methodologies in sustainability discourse analysis, (ii) Corporate-financial logic and environmental ethics, climate justice, and planetary boundaries, and (iii) Critical discourse literacy in Environmental Education for Sustainability (EEfS). Each of these categories are reviewed below in connection with the present study.

The n-gram analysis in NLP has become a foundational method for capturing sequences of *n* words. Amaliyah et al. (2025) show the application of n-gram models with Support Vector Machine (SVM) classifiers to identify patterns of misleading environmental communications in oil and gas industry sustainability reports. Wahyudi & Alamsyah (2025) use NLP to study public discourse on the adoption of green products in emerging economies. Furthermore, using meta-analysis, Asatani et al. (2020) apply multiple NLP techniques to map research trends about sustainability and SDGs. This study relates to the above-mentioned literature as it applies the NLP method, which will serve as a pedagogical tool for developing critical discourse literacy. These methodological contributions provide the analytical approach adopted in this study.

Conti et al. (2025) conduct keyword analyses of the Chinese ESG reports and show how linguistic choices develop particular human-nature relationships and environmental responsibilities. Park et al. (2023) explore the CSR and ESG literature and show a thematic dominance of corporate actors, stakeholder management, and governance frameworks, with limited structural critique and limited engagement with climate justice. Wells et al.'s (2021) analysis shows that corporate discourse considers sustainability as a

matter of brand management and stakeholder relations. This literature review shows that corporate-financial logic dominates contemporary sustainability discourse. Environmental ethics, climate justice frameworks, and planetary boundaries discourse is marginalized in mainstream sustainability communications. Critical discourse analysis examines how this dominance is reflected and reinforced through language.

Critical discourse analysis (CDA) provides an understanding of how language in the academic discourse operate particularly by emphasizing certain understandings of the sustainability concept while marginalizing others. Several studies apply CDA to show how corporate and policy texts have neutralized critique and preserved business-friendly governance conditions (Wells et al., 2021; Küfeoğlu, 2022). Nervino et al. (2025) show, by using corpus-assisted discourse analysis that sustainable finance discourse emphasizes financial returns and risk management over ethical obligations. Asatani et al. (2020) reveal that goals related to economic development receive more research attention compared to those focused on ecological integrity or social justice. Funk et al. (2024) consider SDG achievement and research attention to show that the academic discourse may not effectively translate into sustainability outcomes. Lee et al. (2023) compare ESG discourse in news versus academic literature and find that both emphasize governance and financial dimensions over environmental ethics or social justice. Critical discourse literacy is the capacity to identify, examine, and challenge the assumptions embedded in dominant discourses, and is proposed here as a key pedagogical outcome of EEfS. Given the dominance of corporate-financial logic in the reviewed literature, this study suggests that EEfS should incorporate critical discourse literacy to help learners question and challenge dominant sustainability narratives.

3. Methodology

The study adopted a mixed-methods design structured around an explanatory sequential logic, meaning that quantitative findings were generated first and then used to guide and deepen a subsequent qualitative inquiry. This sequencing was deliberate: the large-scale statistical identification of dominant terms provided the analytical scaffolding upon which discursive interpretation was later built. The corpus comprised 604 scientific articles indexed in Scopus during 2023 and 2024, a period chosen to capture the most recent consolidation of scholarly vocabulary in corporate sustainability research. Restricting the dataset to Scopus-indexed, English-language publications ensured indexing consistency and comparability across sources, though it necessarily excludes contributions published in other languages or in regional journals not covered by that database.

In the first, quantitative phase, the full text of each article was subjected to Natural Language Processing (NLP) techniques using Python. Text cleaning followed a standardized pipeline: lemmatization was applied to reduce each word to its morphological root, and stopwords were removed using customized lists that went beyond generic English function words to also exclude discipline-neutral academic terms unlikely to carry substantive meaning in the sustainability domain. Following this preprocessing, unigrams (single words), bigrams (two-word sequences), and trigrams (three-word sequences) were extracted and counted using the CountVectorizer function. This n-gram approach captures not only isolated concepts but also their most frequent collocations, providing a richer picture of how ideas cluster and co-occur in the literature than frequency counts of individual terms alone could offer.

The second phase introduced qualitative framing analysis, conducted with the assistance of NVivo software. The focus shifted from raw frequency to discursive context: rather than simply noting that a given term appears often, the analysis examined how it was used, in what argumentative position, and alongside what other concepts. This examination concentrated on three structurally strategic sections of each article, specifically the abstract, the introduction, and the conclusions, since these are the spaces where authors most explicitly articulate their conceptual framing and normative orientation. The analytical objective was to characterize the predominant semantic frameworks shaping the field, identifying whether sustainability was positioned as an intrinsic societal imperative or, conversely, as an instrumental variable subordinate to organizational and financial goals.

In the third and final phase, quantitative frequency data and qualitative framing findings were systematically triangulated. This integration step is methodologically essential: frequency distributions reveal what the field talks about most, but they cannot, by themselves, explain what those conversations mean or what values and assumptions underpin them. Triangulation provided the interpretive bridge between the statistical patterns and their discursive significance, ensuring that conclusions rested on mutually reinforcing evidence rather than on either type of analysis alone.

4. Results and discussion

The frequency distributions generated by the n-gram analysis paint a consistent and revealing picture of how corporate sustainability is currently framed in high-impact academic literature. Among unigrams, the term “sustainability” emerged as the most frequent by a wide margin (66,473 occurrences), which is unsurprising given the corpus. What is analytically significant is the company this term keeps: “company” (57,091) and “corporate” (39,380) rank as the second and third most frequent terms, while “financial” (17,262) also appears prominently. This pattern already signals that sustainability, in this body of literature, is not primarily discussed in ecological or ethical terms but rather as a property or attribute of the corporate form.

The bigram frequencies reinforce this reading with considerable force. The most frequent two-word sequences are “social responsibility” (9,035), “corporate social” (8,107), and “corporate sustainability” (6,318), with “financial performance” (4,349) appearing as another high-frequency collocation. The preponderance of “corporate social responsibility” as a conceptual anchor, confirmed by its trigram frequency of 6,795 and thus by far the most common three-word phrase in the corpus, is analytically significant. CSR is a framework historically developed to articulate the voluntary obligations of firms toward society, but its near-ubiquity here points to a disciplinary tendency to conceptualize sustainability almost exclusively through the lens of what it means for and to the firm. Sustainability thus enters academic discourse less as a systemic imperative rooted in thermodynamic limits, biodiversity thresholds, or intergenerational justice, and more as a managerial performance dimension subject to measurement, reporting, and optimization.

Other high-frequency trigrams deepen this interpretation. Sequences such as “environmental social governance” (1,088), “sustainability development goal” (1,061), and “supply chain management” (702) indicate that when broader environmental or global frameworks do appear, they tend to be operationalized within corporate governance and supply chain contexts rather than discussed on their own terms. The presence of “board gender diversity” (643) and “human resource management” (606) among the most frequent trigrams further illustrates how sustainability concerns are increasingly absorbed into organizational management discourse, stripped of their broader normative and political dimensions.

Qualitative framing analysis confirmed and extended these quantitative signals. Across abstracts, introductions, and conclusions, the dominant discursive logic positions sustainability as an instrument for risk mitigation and economic value creation. Terms like “financial performance” and “ESG performance” do not appear as mere neutral descriptors; they function as the evaluative criteria against which sustainability initiatives are measured. This framing forecloses, or at least marginalizes, alternative conceptualizations rooted in environmental ethics, climate justice, or planetary boundaries, none of which featured among the high-frequency terms. The relative absence of vocabulary associated with ecological limits, sufficiency, or redistributive justice is as informative as the presence of corporate and financial terminology.

Taken together, the quantitative and qualitative findings converge on the same conclusion: the dominant academic discourse on corporate sustainability is structured by a corporate-financial logic that subordinates ecological and social imperatives to organizational management objectives. This is not merely a semantic observation. Discourse shapes research agendas, policy recommendations, and educational curricula. When the field systematically frames sustainability as a means to financial ends rather than as an end in itself, it constrains the range of solutions that researchers conceptualize and the criteria by which those solutions are evaluated. For Environmental Education for Sustainability, this has direct consequences: students trained primarily within this hegemonic framing may develop sophisticated technical skills in ESG reporting or sustainability metrics while remaining largely unequipped to question the structural logics of growth, extraction, and capital accumulation within which those metrics are embedded.

It is important to note that these results do not suggest that corporate engagement with sustainability is inherently misguided, nor that the body of literature analyzed lacks scholarly merit. Rather, they highlight the value of critical discourse literacy as a complement to technical competence in sustainability education. Recognizing the assumptions built into dominant frameworks is a precondition for transcending them.

5. Conclusion

This study examines the academic discourse on corporate sustainability and its implications for Environmental Education for Sustainability (EEfS). By using a mixed-methods design combined with n-gram and qualitative framing analysis, this study identifies a consistent and structurally significant pattern. The study finds that the vocabulary in corporate sustainability literature is organized around a

corporate-financial logic. This logic is more relevant to organizational performance objectives than it is to ecological limits and ethical principles.

The n-gram analysis of a corpus of 604 Scopus-indexed articles included the frequency patterns of "corporate social responsibility," "financial performance," and "ESG performance." The frequency patterns reveal a disciplinary tendency to consider sustainability as a managerial variable. Furthermore, the qualitative framing analysis confirmed that this vocabulary is not neutral but is instead used as an instrument of risk management rather than for ecological and intergenerational justice. These findings highlight the need to foster critical discourse literacy as a core pedagogical outcome. Critical discourse literacy enables learners to operate within dominant frameworks while challenging the underlying values and assumptions through environmental ethics and climate justice.

Future research will incorporate a broader corpus beyond Scopus-indexed, English-language publications to examine whether the academic discourse on corporate-financial logic is applied uniformly or more predominantly in developed countries and how this discourse has changed over time. Locational research through a comparative study of academic discourse across regions including Latin America, the Global South, and East Asia may provide insights into local ecological vulnerabilities, governance, and cultural values. Such a locational study will provide local pedagogical implications and place-based frameworks for EEfS. A longitudinal analysis will compare discourse patterns across different decades. The ultimate goal is to close the gap between the principles of EEfS and the underlying assumptions of corporate sustainability through sustained efforts from researchers, educators, and policymakers.

References

- Amaliyah, F., Putri, A. H., Andyna, N. P., & Amri, Z. K. (2025). PRISMA systematic review: The application of natural language processing (NLP) to identify greenwashing in sustainability reports within the oil and gas industry. *Journal of Innovation Materials, Energy, and Sustainable Engineering*, 3(1), 58-74.
- Asatani, K., Takeda, H., Yamano, H., & Sakata, I. (2020). Scientific Attention to Sustainability and SDGs: Meta-Analysis of Academic Papers. *Energies*, 13(4), 975.
- Conti, S., Locatelli, L., Brombal, D., & Szeto, P. Y. (2025). The language of ESG sustainability reports of Chinese listed companies: A preliminary corpus-assisted keyword analysis from an ecolinguistic perspective. *Chinese Language and Discourse*.
- Funk, C., Tönjes, E., Teuber, R., & Breuer, L. (2024). Reading between the lines: The intersection of research attention and sustainable development goals. *Sustainable Development*, 32(5), 4545-4566.
- Huckle, J., & Sterling, S. R. (Eds.). (1996). *Education for sustainability*. London: Earthscan from Routledge.
- Küfeoğlu, S. (2022). Emerging Technologies. In *Emerging Technologies. Sustainable Development Goals Series*. Springer, Cham. https://doi.org/10.1007/978-3-031-07127-0_2.
- Lee, H., Lee, S., Lee, K., & Jang, K. (2023). ESG Discourse Analysis Through BERTopic: Comparing News Articles and Academic Papers. *CMC-Computers Materials & Continua*. 75(3), 6023.
- Nasibulina, A. (2015). Education for sustainable development and environmental ethics. *Procedia-Social and Behavioral Sciences*, 214, 1077-1082.
- Nervino, E., Lan, G., & Cheung, J. O. (2025). Sustainability as a business opportunity: a corpus-assisted discourse analysis of sustainable finance discourse. *Applied Linguistics Review*, 16(5), 1953-1982.
- Park, J. G., Park, K., Noh, H., & Kim, Y. G. (2023). Characterization of CSR, ESG, and Corporate Citizenship through a Text Mining-Based Review of Literature. *Sustainability*, 15(3), 3892.
- Putilova, E., Tsiplakova, Y., Diachkova, A., & Knysh, E. (2023). Environmental education and its principles. *E3S Web of Conferences*, 431, 09003. EDP Sciences.
- Tilbury, D. (1995). Environmental education for sustainability: Defining the new focus of environmental education in the 1990s. *Environmental education research*, 1(2), 195-212.
- United Nations. (2016). *The sustainable development goals 2016*. Retrieved from <https://sdgs.un.org/goals> (accessed on 24 February 2026).
- UNESCO. (2020). *Education for sustainable development: A roadmap*. Paris: UNESCO.
- Wahyudi, R., & Alamsyah, A. (2025). Leveraging transformer-based NLP to analyze public discourse on green product adoption in emerging economies. In *Proceedings of the 2025 International Conference on Communication, Computing, Networking, and Control in Cyber-Physical Systems (CCNCPS)* (pp. 385–390). IEEE. <https://doi.org/10.1109/ccncps66785.2025.11135887>
- Wells, V., Athwal, N., Nervino, E., & Carrigan, M. (2021). How legitimate are the environmental sustainability claims of luxury conglomerates. *Journal of Fashion Marketing and Management*, 25(4), 697-722.