A Multimodal Discourse Analysis of Environmental Narratives in Alibaba's ESG Communications on Its Official Website

Donglan Wen*

School of Jiangxi Normal University, Nanchang 330000, Jiangxi, China

*Corresponding author: Donglan Wen, 410578712@qq.com

Abstract: Based on the theory of multimodal discourse analysis, this study analyzes the synergistic effects of textual, visual, and other semiotic resources in Alibaba's official ESG reports, revealing how these resources construct environmental narratives and empower corporate image through semiotic practices. The research, conducted from the dimensions of representational meaning, interactive meaning, compositional meaning, and modal synergy, deciphers how the semiotic system deeply integrates ecological concepts. The findings indicate a high degree of text-image synergy during the 2022–2024 period, effectively presenting themes such as "Repairing the Green Planet," "Digital Empowerment for Carbon Reduction," and "Clean Energy and Circular Economy." However, the visual presentation in the 2025 report is relatively concentrated on flowcharts and data graphs, showing insufficient multidimensional representation of environmental modules and public engagement, with an overall weakening of text-image synergy. This study enriches the application of multimodal analysis in the field of environmental narrative research. Its theoretical and practical value lies in revealing the strategic role of semiotic synergy in building a sustainable brand and providing a communication pathway reference for promoting public environmental awareness and sustainable consumption behaviors.

Keywords: Multimodal discourse analysis; Environmental narrative; Alibaba official website; ESG report; Semiotics

1. Research background

With the increasing severity of global climate change and environmental issues, public environmental awareness has significantly strengthened. The concept of green environmental protection has become a topic of widespread concern across society. Among these, corporate green initiatives are a focal point of discussion. This concept, centered on environmental protection and sustainable development, aims to meet consumer needs through business activities while safeguarding the natural environment and social interests. In response to this societal concern, companies engage in multi-dimensional environmental storytelling, with

the mainstream paradigm being the construction of environmental narratives through ESG (Environmental, Social, and Governance) reporting. Consequently, ESG has become a core narrative framework for corporate strategic sustainability. In contemporary business society, an increasing number of enterprises use ESG reports to narrate their ethical stories and shape a sustainable development image. The "E" (environmental) component in ESG, representing environmental responsibility, attracts the highest public attention. Many companies are keen on packaging their image with terms like "green," "sustainable," and "carbon neutral," signaling "we are responsible" to consumers through various environmental displays. ESG has evolved from a frontier theory into an international assessment system with clear indicators. Companies that consistently achieve leading ratings become exemplary models. Taking Alibaba Group as an example, it has long maintained an "A" rating in the authoritative MSCI ESG ratings, signifying its broad recognition in the global capital market for its governance structure, environmental management, and social responsibility. This study uses multimodal discourse analysis to focus on how the environmental aspects of Alibaba's ESG reports reflect environmental narratives, deeply understanding how Alibaba embodies environmental values in its ESG reports, thereby revealing its sense of responsibility among global technology companies and providing useful references for other enterprises practicing ESG principles.

2. Literature review

Multimodal Discourse Analysis (MDA) is an important interdisciplinary development between linguistics and social semiotics since the 1990s. Traditional discourse analysis primarily focuses on meaning construction at the linguistic level, whereas MDA posits that meaning generation relies not only on language but also on various semiotic modes such as images, sound, color, layout, gesture, and spatial arrangement [1]. In other words, social meaning is constructed through multiple modes working together. Kress and van Leeuwen, in Reading Images: The Grammar of Visual Design, proposed the "visual grammar" theory, which became a foundational study for multimodal analysis [1]. The application areas of MDA are extensive. For instance, Jewitt and Tang used MDA to examine how language, images, charts, and layouts in textbooks, classroom interactions, and multimedia courseware collectively construct knowledge and influence the learning process [2]. Ledin and Machin analyzed how multimodal resources in news, advertisements, and political propaganda collaboratively construct ideology, power relations, and identity [3]. Adami and Zappavigna analyzed the multimodal features of digital texts such as websites, social media (e.g., TikTok, Instagram), memes, and video games, focusing on their interactivity, hypertextuality, and dynamism [4,5]. It is evident that past research on multimodality has largely focused on texts and images in education, news, and advertising. In recent years, MDA has gradually extended into the fields of Corporate Social Responsibility (CSR) and Environmental Communication. Companies use annual reports, sustainability reports, ESG reports, and promotional materials to tell stories of "green innovation" and "sustainable development" [6]. However, much of this research focuses on the linguistic analysis of ESG texts, paying insufficient attention to the construction of meaning at the visual level. By conducting a multimodal discourse analysis of the text and images in corporate ESG reports, researchers can not only gain a deeper understanding of how companies shape their environmental image through the synergy of different semiotic resources like text and images but also reveal how these semiotic systems, through complementary and reinforcing methods, jointly construct a powerful environmental narrative and foster public identification with and support for corporate environmental commitments.

3. Theoretical framework

Multimodal discourse analysis is based on Halliday's Systemic Functional Linguistics (SFL). SFL proposes that language has three meta-functions: ideational, interpersonal, and textual. Kress and van Leeuwen extended this framework to multimodal domains such as visuals, color, and layout, forming a visual grammar analytical framework. (As shown in **Figure 1**) This framework emphasizes that signs like images, colors, and text do not exist in isolation but convey complex information collectively through "representational meaning," "interactive meaning," and "compositional meaning." The integration of these modalities constitutes multimodal semiotic resources and generates meaning to achieve effective communicative purposes [7].
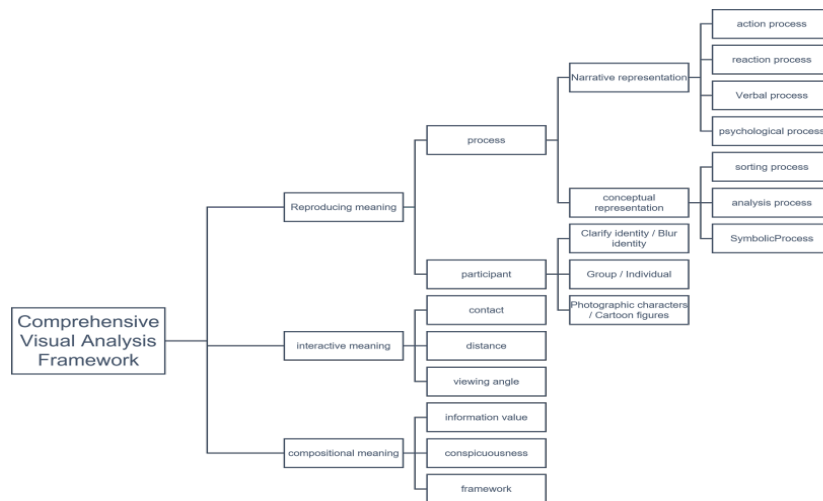
Figure 1. Comprehensive visual analysis framework

3.1. Representational meaning

Representational meaning corresponds to the ideational function in SFL, revealing how images represent "what the world is like" and "social action" [8]. Kress and van Leeuwen divided it into two categories: narrative representation and conceptual representation. Narrative representation depicts dynamic processes and actions, often containing elements like vectors (action lines), actors, and goals within the image, expressing "what is happening." Conceptual representation portrays static, classificatory, or symbolic relationships, embodying the meaning of "what something is," such as charts, logos, or conceptual displays.

3.2. Interactive meaning

Interactive meaning corresponds to the interpersonal function in SFL, focusing on how images establish relationships with the viewer. Kress and van Leeuwen proposed three core dimensions: gaze, social distance, and perspective. Gaze indicates whether figures in the image make eye contact with the viewer, forming a "demand" relationship; if not, it is an "offer" relationship. Social distance is conveyed through shot scale (e.g., long shot, close-up), representing social intimacy or detachment—a long shot signifies objectivity and distance, while a close-up conveys intimacy and involvement. Perspective involves camera angle: a high angle (looking down) can express power or authority, a low angle (looking up) can convey respect or equality, and an eye-level angle suggests reciprocity and empathy.

3.3. Compositional meaning

Compositional meaning corresponds to the textual function in language, concerned with how various elements in an image are organized into a coherent whole. Its analytical dimensions include: 1) Information Value: Different zones carry different informational functions. Generally, the left represents the "Given" (known information), the right the "New"; the top embodies the "Ideal," the bottom the "Real"; central elements signify core information, while marginal elements provide supplementary details. 2) Salience: Highlighting visual focus through size, color, contrast, focus, etc. 3) Framing: Distinguishing or connecting different information units through borders, whitespace, or dividing structures. The compositional structure not only determines the flow of visual attention but also reflects the organizational logic of ideology [9].

4. Research design

4.1. Research questions

This study aims to employ multimodal discourse analysis from a social semiotic perspective, utilizing the visual grammar theory proposed by Kress and van Leeuwen, to explore how visual signs in Alibaba's ESG reports participate in constructing the company's environmental narrative and showcasing its sustainable development ideal. The research will revolve around the following two core questions:

(1) How does Alibaba's ESG report visually embody environmental narratives and shape a green, sustainable corporate image through representational meaning, interactive meaning, and compositional meaning?

(2) How does Alibaba's ESG report construct environmental narratives through text-image synergy mechanisms?

4.2. Research object

This study takes the ESG reports released by Alibaba Group from 2022 to 2025 as the visual text corpus. The reports contain various semiotic modalities, including textual narratives, data charts, images, color design, and layout, possessing typical multimodal characteristics.

4.3. Environmental theme keyword analysis

To identify the environmental discourse themes embodied in Alibaba's ESG reports, this study employs a micro word cloud for Chinese text frequency analysis. By importing the environmental sections of the four ESG reports, statistics were gathered on high-frequency words. Combined with the textual context of the reports, the core issues of the corporate environmental narrative were extracted. The analysis focused on vocabulary related to environmental dimensions in the reports, such as energy utilization, carbon emissions, waste management, water resource protection, and green technology application. The results of the thematic keyword extraction are shown in **Table 1**.

Table 1. Environmental theme keywords in Alibaba ESG reports (2022–2025)

Year	Core Theme	Theme Keywords	Semantic Explanation and Research Findings
2022	Repairing the Green Planet & Carbon Neutrality Commitment	Carbon neutrality, Green campus, Data center, Energy transition, Clean energy, Shared governance	Established the "Repairing the Green Planet" concept, proposed carbon neutrality goals, focused on internal energy saving, green office, and campus energy efficiency, laying the visionary framework.
2023	Digital Empowerment & Ecological Carbon Reduction	Digitalization, Platform ecosystem, Scope 3+ (suppliers, service providers, consumers), Low-carbon logistics, Circular economy	Narratively shifted towards "platform empowerment" through digital technology promoting supply chain carbon reduction, moving from "self-emission reduction" to "ecological collaboration."
2024	Repairing the Green Planet & Deepening Circular Economy	Carbon neutrality, Circular packaging, Green logistics, 4R circular economy, Ecological protection	Showcased achievements driven by both "digital technology + clean energy," forming a green operational system, with narrative logic shifting towards "practice deepening."
2025	Climate Governance & Standardized Disclosure	Climate risk governance, Carbon disclosure, 4R circular economy, Digital circular logistics, Biodiversity	Entered an institutionalization and internationalization phase, focusing on climate governance and disclosure, marking the maturity of ESG discourse and alignment with global standards.

4.4. Analytical summary

From 2022 to 2025, the environmental narrative in Alibaba's ESG reports demonstrates a systematic evolution from "carbon neutrality commitment" to "digital empowerment for carbon reduction," then to "clean energy practice" and "governance disclosure." Although high-frequency words are similar across years, their semantic functions gradually evolve. The emphasis differs each year: early years focused on strategic vision and system building, mid-term shifted towards technology and ecological collaboration, and later years concentrated on institutionalization and international disclosure. This reflects the digitalized, ecologized, and governance-oriented characteristics of corporate environmental discourse, clearly illustrating the company's strategic adjustments during development, adapting to the requirements of the times and building a sustainable corporate development image.

5. Multimodal construction of environmental narratives in Alibaba's ESG reports

Based on the three meta-functions theory of Halliday's Systemic Functional Linguistics, this paper conducts a multimodal discourse analysis of Alibaba Group's ESG reports from 2022 to 2025 (all figures are from Alibaba ESG website). It primarily reveals the environmental narrative logic and visual expression strategies at different stages through representational meaning, interactive meaning, and compositional meaning, thereby understanding how digital platform companies construct a sustainable development corporate image through multimodal discourse.

5.1. Representational meaning

Figure 2 constructs a typical narrative structure through the action chain of "user — autonomous delivery vehicle — package," representing the high-frequency thematic logic of digitalization and Scope 3+ user participation in Alibaba's 2023 ESG environmental narrative. The smart delivery vehicle and the user's package pickup action in the image together form a visualized node in the logistics process, reflecting the low-carbon value generated by the synergy of "technology-user-platform." Digital and intelligent circular logistics can improve delivery services more efficiently, better achieving ecological carbon reduction.

Figure 3 presents a scene of "a family cycling in a natural environment," depicting low-carbon travel and green lifestyle as a stable, sustainable life tableau. This belongs to a composite representational method combining narrative and conceptual structures. The natural greenery, as a symbolic attribute, reinforces key issues highlighted in the 2024 ESG report, such as the 4R circular economy, natural ecological protection, and green lifestyle advocacy. The characters' actions and the portrayal of families jointly practicing green ecological ideals demonstrate that action for green carbon reduction should be taken from individuals to enterprises. The representational meanings of the two images correspond respectively to the company's narrative shift from "technology-driven emission reduction" to "lifestyle-driven emission reduction," jointly constituting the multimodal construction of ESG environmental narratives.

Figure 2. Alibaba ESG 2023
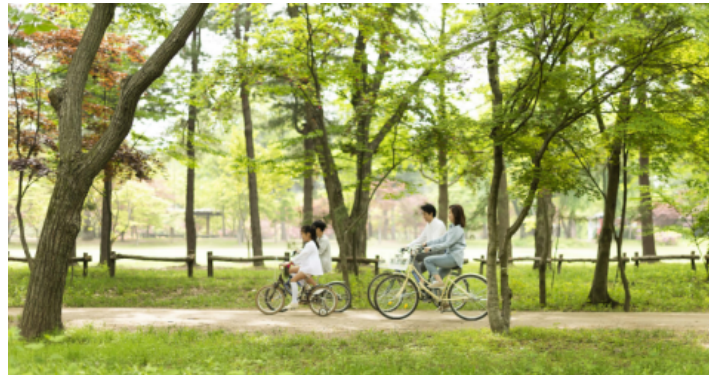
Figure 3. Alibaba ESG 2024

5.2. Interactive meaning

Figure 4 showcases many hands planting trees, representing public participation and collective action. Multiple hands of different skin tones and ages jointly hold a small sapling with soil. The circular composition forms a "circle of unity," presenting a cooperative action scene. It belongs to the offer type image, displaying a social relationship of "collective action." The close-up shot of the hands creates a high sense of intimacy, implying equal participation and shared responsibility. This image highly aligns with the "Carbon Reduction Friendly Actions" section of the report, emphasizing the construction of public participation and cooperative ecology. The hands forming a circular composition visually reconstruct the core concept of "Repairing the Green Planet" – every actor (consumer, employee, enterprise) is jointly supporting the future of the planet.

Figure 4. Alibaba ESG 2022

Figure 5 uses a close-up side shot, showing a child focused on depositing a used cardboard box into a "Cainiao" recycling bin. It is a typical "offer" image. The viewer is placed in the position of an observer, witnessing and learning from this environmental practice. **Figure 4** highlights "who governs nature together," while **Figure 5** highlights "how everyone acts in daily life." Together, they constitute the multimodal environmental narrative chain in Alibaba's ESG report, from presenting "co-governance subjects" to

Figure 5. Alibaba ESG 2023

"daily practice paradigms." This also reflects the main thread of the 2023 theme, Scope 3+, which expresses

environmental narrative through consumers as participants.

    Figure 6 constructs the relationship between image participants and viewers through various visual strategies. First, at the level of visual contact, figures in the image mostly do not establish direct eye contact with the lens; they face the scene or gesture sideways. This "offer" type gaze does not issue an interactive demand to the viewer but presents the event itself as an object for observation, thereby positioning the viewer as an observer and witness of a public event. Secondly, in the construction of social distance, the image uses medium-to-long shots, allowing the complete presentation of the figures' overall appearance and their surrounding spatial environment. This distance conforms to the visual conventions of formal public occasions, weakening the interference of personal emotions and emphasizing the institutional identity and collective attributes of the figures, effectively constructing a discursive space centered on officialdom, publicity, and international cooperation.

Figure 6. Alibaba ESG 2024

    Figure 7 constructs a highly engaging viewing relationship through de-faced bodily presentation and intimate-distance composition. The image shows only multiple hands surrounding green plants, without establishing direct eye contact, belonging to an "offer" type interaction, prompting viewers to focus their attention on the ecological theme itself. In terms of social distance, the close-up composition shortens the psychological distance between the viewer and the image, enhancing emotional involvement. A high-angle perspective gives the viewer a reflective position to examine the overall human-nature relationship, while the high modality value

Figure 7. Alibaba ESG 2025

formed by high-saturation green and clear details further strengthens the image's sense of reality and moral appeal, transforming the viewer from an observer into a potential responsible subject.

5.3. Compositional meaning

Figure 8 shows a symbolic distribution with a left-right contrast structure. The left-right distribution represents the relationship between Given and New information, which can be described as a variation of the Ideal-Real (top-bottom) relationship. The left side features a large white recycling symbol occupying the visual subject, constituting the "Ideal" zone. The right side shows a pile of discarded electronic devices, forming the "Real" zone. This left-right structure embodies a dialogue between the ideal (circular economy) and reality (e-waste). Visually, the image constructs a value transformation path from "waste to circular," corresponding to "Green Campus and Resource Recycling" in Alibaba's 2022 ESG report. This vividly reappears the process from electronic waste to green cycle, constructing a visual narrative framework of "sustainable production and consumption," and also expresses the themes of repairing the green planet and carbon reduction for 2022.

Figure 9 shows the hands of multiple participants surrounding plant identification signs. Through a center–periphery information structure, it clearly presents a visual logic centered on "biodiversity knowledge as the core, with the public as peripheral supporters and participants." Through its visual structure, the image constructs the symbolic meaning of "multiple actors collaborating around an environmental core," reinforcing the platform-driven, networked, and participatory environmental narrative emphasized in the 2023 theme.

Figure 10 is an overhead view of the park, presenting a typical "ideal above, reality below" layout in terms of compositional meaning. This achieves a visual connection between the company's green vision and real-life scenes. The upper part features streamlined curves formed by the building canopy, resembling a "technology ring," conveying a vision for future cities and green technology at the ideal level. The lower part consists of real natural landscapes (grass, trees, water, and stone paths) as sustainable living scenes at the realistic level. Through the integration of architectural and natural forms, a multidimensional narrative of "humans actively adapting to nature and coexisting harmoniously with the ecosystem" is constructed, conveying the ecological philosophy of sustainable development. This aligns well with the ESG theme of Alibaba in 2025.

| Figure 8. Alibaba ESG2022 | Figure 9. Alibaba ESG 2023 | Figure 10. Alibaba ESG2025 |

5.4. Text-image synergy analysis

In terms of interactive meaning, the ESG report constructs a multi-layered relational structure ranging from "co-governance subjects" to "everyday actors" through the integrated use of gaze, shot scale, and perspective. On the one hand, images such as group photos and portraits present enterprises, social organizations, and the public from eye-level perspectives and medium-close shots, symbolizing collaborative governance and equal partnership. On the other hand, close-up shots like hands planting trees guide viewers to immerse themselves in specific environmental practices through intimate social distance and offering images, reinforcing actionable and replicable participation models for individuals. Images of pandas and natural landscapes, by maintaining appropriate distance and employing eye-level perspectives, convey respect for nature and the concept of coexistence, infusing the company's environmental commitments with emotional warmth and moral legitimacy.

Notably, there is a weakening of multimodal synergy in the 2025 report. The environmental section of Alibaba's 2025 ESG report has clearly entered a governance-oriented phase focused on "climate governance and standardized disclosure," with high-frequency terms concentrated on institutionalized and standardized concepts such as "climate risk governance," "carbon disclosure," "4R circular economy," and "digital circular logistics." This indicates that the company's environmental discourse has further advanced from "commitment–practice" to "governance–disclosure." However, at the multimodal level, the report extensively employs flowcharts, indicator charts, and structural diagrams as primary visual resources, significantly reducing the representation of public action scenarios, natural environmental imagery, and multi-stakeholder interactions.

This weakening of multimodal synergy reveals that as ESG discourse transitions from "storytelling" to rules and data, a lack of corresponding multimodal design approaches can lead to governance logic

overshadowing narrative space. While the textual level becomes increasingly mature and institutionalized, the visual level struggles to sustain public engagement and motivation for action. This not only limits the communicative power of the environmental narrative but also, to some extent, undermines the construction of the company's "perceived sense of responsibility." Therefore, further efforts are needed in the design of ESG reports to better align with themes and foster interaction with viewers.

6. Research conclusion

This study adopts a multimodal discourse analysis framework from a social semiotic perspective to systematically examine the textual and visual resources in the environmental sections of Alibaba's ESG reports from 2022 to 2025. The findings reveal that, in terms of representational meaning, Alibaba's environmental narrative has shifted from a technology-driven emission reduction focus to a lifestyle-oriented approach. Early images predominantly featured "technology-infrastructure" scenarios such as data centers, green campuses, and smart logistics, emphasizing the emission reduction effects of digital technologies and platform capabilities. Subsequently, the reports gradually incorporated everyday life scenes, such as family cycling and public participation in recycling, making green lifestyles a key visual theme. This evolution marks a narrative upgrade from corporate-led actions to multi-stakeholder engagement.

However, this progression was not consistently synchronized between text and images, particularly in the 2025 report, where textual advancement was accompanied by a noticeable decline in visual support, reflecting a weakening of multimodal synergy.

Disclosure statement

The author declares no conflict of interest.

References

[1]   Kress G, van Leeuwen T, 2006, Reading Images: The Grammar of Visual Design (2nd ed.). Routledge, London.

[2]   Jewitt C, 2009, The Routledge Handbook of Multimodal Analysis (Chapter on Education). Routledge, London.

[3]   Ledin P, Machin D, 2018, Doing Visual Analysis: From Theory to Practice. Sage, New York.

[4]   Adami E, 2017, Multimodality, in The Routledge Handbook of Language and Media. Routledge, London, 149–164.

[5]   Zappavigna M, 2016, Social Media Photography: Construing Subjectivity in Instagram Images. Visual Communication, 15(3): 271–292.

[6]   Cheng R, Zhang D, 2017, The Current State, Characteristics, and Development Trends of Multimodal Discourse Analysis Research in China: Based on Journal Publications. Foreign Languages in China, 14(3): 40–43.

[7]   Zhang D, 2009, Exploring a Comprehensive Theoretical Framework for Multimodal Discourse Analysis. Foreign Languages in China, 2009(1): 24–30.

[8]   Li Z, 2003, A Social Semiotic Analysis of Multimodal Discourse. Foreign Language Research, 2003(5): 1–7.

[9]   Zhang D, 2025, A Study on the Applicability of Social Semiotics in Multimodal Discourse Construction. Journal of Beijing International Studies University, 47(2): 19–28.