Published: 2021-01-29

AI Based Analysis of Environmental Disclosure Tone and Readability

Corbin Foster, Wesley Rivera, Ellie Greene

Abstract

This research introduces a novel computational framework for analyzing corporate environmental disclosures through the dual lenses of linguistic tone and textual readability, employing artificial intelligence techniques that diverge from traditional sentiment analysis and financial text mining approaches. We propose a hybrid methodology combining transformer-based language models with psycholinguistic dictionaries and graph-based coherence metrics to quantify not just what is said about environmental performance, but how it is communicated. Our approach uniquely integrates three unconventional dimensions: (1) a multi-scale tone analysis capturing micro-sentiment fluctuations alongside macro-narrative arcs, (2) a readability assessment that accounts for domain-specific environmental terminology rather than general linguistic complexity, and (3) a coherence metric evaluating logical flow between environmental claims and supporting data. We apply this framework to a corpus of 2,500 environmental disclosures from SP 500 companies between 1995 and 2004, revealing previously undocumented patterns of strategic obfuscation in high-risk industries. Results demonstrate that companies in environmentally sensitive sectors employ significantly more complex sentence structures when discussing negative environmental impacts compared to positive achievements (p ¡ 0.01), while maintaining an artificially optimistic tone through selective positive framing. Furthermore, we identify a 'readability gap' where environmental performance metrics are presented with 42

Keywords: environmental disclosure, computational linguistics, readability analysis, tone analysis, corporate communication, artificial intelligence, text mining

1 Introduction

Corporate environmental disclosure has emerged as a critical domain of organizational communication, bridging the informational needs of stakeholders with corporate accountability mechanisms. Traditional approaches to analyzing these disclosures have predominantly fo-

cused on content quantification—measuring the presence or absence of specific environmental topics, performance metrics, or compliance statements. However, this content-centric paradigm overlooks the nuanced linguistic dimensions through which environmental information is strategically communicated. The manner in which environmental performance is described—the tone, readability, and narrative coherence—may convey as much strategic intent as the factual content itself. This research addresses this gap by developing and applying an artificial intelligence framework specifically designed to analyze the stylistic and rhetorical dimensions of environmental disclosures.

Our investigation is motivated by three research questions that have received limited attention in the environmental accounting and computational linguistics literature. First, how do corporations strategically modulate linguistic tone when discussing environmentally sensitive topics compared to neutral or positive environmental achievements? Second, to what extent do companies manipulate the readability of environmental disclosures, potentially creating barriers to comprehension for non-expert stakeholders? Third, how coherent are the narratives connecting environmental claims with supporting evidence, and does this coherence vary systematically across industries with different environmental risk profiles? These questions probe beneath the surface content of disclosures to examine the linguistic architecture of environmental communication.

Previous research in financial text analysis has established that linguistic tone influences investor perceptions and market reactions. However, these approaches have typically employed general-purpose sentiment dictionaries that fail to capture the domain-specific linguistic nuances of environmental discourse. Similarly, readability research has relied on formulas like the Flesch-Kincaid index, which were developed for general educational texts and may misrepresent the complexity of technical environmental communication. Our research contributes by developing domain-adapted measures that account for the specialized lexicon and syntactic patterns characteristic of environmental reporting.

The theoretical foundation of this work integrates insights from impression management

theory, which suggests organizations strategically craft disclosures to shape stakeholder perceptions, with computational linguistics approaches that enable large-scale textual analysis. We posit that environmental disclosures represent a particularly rich domain for studying strategic communication because they sit at the intersection of regulatory requirements, public relations considerations, and substantive performance reporting. The period from 1995 to 2004 provides an ideal context for this analysis, as it encompasses both the emergence of voluntary environmental reporting frameworks and increasing regulatory attention to corporate environmental transparency, yet precedes the widespread standardization of sustainability reporting that occurred in subsequent years.

2 Methodology

Our methodological approach represents a departure from conventional text analysis in environmental research through its integration of three innovative analytical dimensions within a unified computational framework. The foundation of our analysis is a corpus of 2,500 environmental disclosures extracted from annual reports, sustainability reports, and 10-K filings of SP 500 companies spanning the decade from 1995 to 2004. This temporal scope was selected to capture the evolution of environmental reporting practices during a period of increasing stakeholder attention to corporate environmental performance, while ensuring all documents predate the widespread adoption of standardized sustainability reporting frameworks that might homogenize linguistic patterns.

The first analytical dimension addresses tone analysis through a multi-scale approach that combines dictionary-based methods with machine learning techniques. Rather than employing general sentiment lexicons, we developed a domain-specific environmental tone dictionary through an iterative process of manual annotation and statistical validation. This dictionary captures not only positive and negative valence but also distinct emotional tones relevant to environmental discourse, including concern, optimism, certainty, and mitiga-

tion. At the micro-scale, we analyze tone at the sentence level to capture local fluctuations in emotional expression. At the macro-scale, we employ a novel narrative arc detection algorithm that identifies patterns of tone progression throughout entire documents, distinguishing between linear positive narratives, redemption narratives (negative to positive), and obfuscation patterns (neutralizing negative content within predominantly positive sections).

The second dimension focuses on readability assessment through a hybrid approach that modifies traditional readability formulas to account for domain-specific complexity. Standard readability measures like the Gunning Fog Index and SMOG formula overweight sentence length and syllable count while underweighting conceptual complexity and domain-specific terminology. Our Environmental Readability Index (ERI) incorporates four weighted components: syntactic complexity (adapted from traditional measures), conceptual density (ratio of environmental technical terms to general vocabulary), inter-sentence connectivity (measured through referential cohesion), and jargon concentration (frequency of industry-specific terminology without contextual explanation). The ERI was validated through a comprehension study with 120 participants reading environmental disclosures and answering factual and interpretive questions about the content.

The third dimension introduces a coherence metric that evaluates the logical flow between environmental claims and supporting evidence. Using a graph-based representation where nodes represent factual claims or data points and edges represent logical connections, we compute coherence scores based on the density of connections between claims and evidence, the presence of logical fallacies (identified through pattern matching), and the consistency of quantitative references throughout the document. This approach moves beyond surface-level cohesion to assess the substantive logical architecture of environmental arguments.

Our analytical framework implements these three dimensions through a pipeline of natural language processing techniques. Documents undergo preprocessing including sentence segmentation, part-of-speech tagging, and named entity recognition with special attention to environmental entities (pollutants, regulations, initiatives). The tone analysis module

employs both our custom dictionary and a support vector machine classifier trained on manually annotated environmental sentences. The readability module computes the ERI through a combination of rule-based and statistical approaches. The coherence module constructs document graphs using dependency parsing and rhetorical structure theory. All analyses were implemented in Python using the Natural Language Toolkit and custom-developed algorithms, with statistical validation of inter-coder reliability for manual annotation components exceeding Cohen's kappa of 0.85.

3 Results

The application of our analytical framework to the corpus of environmental disclosures revealed systematic patterns in tone manipulation, readability variation, and narrative coherence that differed markedly across industries and environmental risk profiles. Our findings challenge the assumption that environmental disclosures represent straightforward factual reporting and instead suggest sophisticated linguistic strategies that may obscure as much as they reveal.

Tone analysis demonstrated that companies in environmentally sensitive industries (chemicals, energy, manufacturing) exhibited significantly different tonal patterns than companies in low-impact sectors (technology, services, finance). High-environmental-impact companies maintained an overall positive tone (mean positivity score of 0.62 on a -1 to +1 scale) despite reporting more negative environmental incidents. This apparent contradiction was resolved through micro-analysis revealing strategic tonal placement: negative information was consistently embedded within sentences that began or ended with positive framing, a pattern we term 'tonal sandwiching.' Furthermore, these companies employed significantly more certainty markers (words like 'definitely,' 'unquestionably,' 'assuredly') when discussing environmental compliance than when discussing performance improvements ($t=4.32$, p¡0.001), suggesting an effort to bolster perceptions of regulatory adherence despite potentially am-

biguous performance records.

Readability assessment revealed a systematic 'complexity gradient' where sections discussing environmental liabilities, regulatory violations, or substandard performance metrics were written with significantly greater syntactic and conceptual complexity than sections highlighting environmental achievements or community initiatives. The mean Environmental Readability Index score for negative performance sections was 14.2 (equivalent to college graduate level comprehension), compared to 9.8 for positive sections (high school level comprehension). This difference was most pronounced in the energy sector, where the complexity gap reached 5.1 ERI points. Further analysis revealed that this complexity derived not only from sentence structure but from strategic use of technical jargon without adequate explanation, creating potential comprehension barriers for non-expert stakeholders.

Perhaps our most striking finding emerged from the coherence analysis, which revealed systematic disconnects between environmental claims and supporting evidence. Across the corpus, 68

Temporal analysis across the 1995-2004 period revealed an intriguing trend: as voluntary environmental reporting became more common, the sophistication of linguistic strategies appeared to increase. Early disclosures (1995-1998) showed more direct correlation between environmental performance and linguistic characteristics, while later disclosures (1999-2004) exhibited greater decoupling, with companies achieving more consistent positive tone regardless of actual performance metrics. This suggests the development of professionalized environmental communication practices that may prioritize perception management over transparent disclosure.

Industry-level comparisons revealed that the extractive industries (mining, oil and gas) employed the most pronounced linguistic strategies, with the highest levels of tonal manipulation and evidence dispersion. In contrast, consumer goods companies showed more straightforward correlations between environmental performance and disclosure characteristics. These industry patterns persisted even when controlling for company size, profitability,

and regulatory exposure, suggesting industry-specific norms in environmental communication.

4 Conclusion

This research makes both methodological and substantive contributions to the understanding of corporate environmental communication. Methodologically, we have developed and validated a novel analytical framework that moves beyond content analysis to examine the linguistic architecture of environmental disclosures. Our integration of multi-scale tone analysis, domain-adapted readability assessment, and graph-based coherence evaluation provides a more nuanced understanding of how environmental information is strategically presented. The Environmental Readability Index represents a significant advance over traditional readability formulas for specialized technical domains, while our tone analysis approach captures the emotional complexity of environmental discourse more effectively than general sentiment analysis tools.

Substantively, our findings reveal systematic linguistic strategies in environmental reporting that have important implications for theory and practice. The consistent patterns of tonal manipulation, strategic complexity, and evidence dispersion suggest that environmental disclosures are crafted with considerable attention to perceptual impact rather than serving as transparent windows into corporate environmental performance. These practices align with impression management theory but extend it by demonstrating how linguistic techniques are deployed in the specific context of environmental accountability.

The identification of industry-specific patterns in environmental communication suggests that normative pressures within industrial sectors may shape disclosure practices as significantly as regulatory requirements. The increasing sophistication of linguistic strategies over time further indicates the professionalization of environmental communication, potentially creating a growing gap between the apparent transparency of disclosures and their actual

comprehensibility and verifiability.

These findings have practical implications for multiple stakeholders. For regulators, they suggest the need for standards addressing not only what environmental information must be disclosed but how it should be presented to ensure comprehensibility and logical coherence. For investors and analysts, they highlight the importance of reading environmental disclosures with attention to linguistic patterns that may signal obfuscation or selective framing. For corporate managers, they raise ethical questions about the appropriate balance between positive framing and transparent communication of environmental challenges.

Several limitations of the current research suggest directions for future work. Our analysis focused on large publicly-traded companies in the United States; extending this framework to smaller companies, international contexts, or different regulatory environments would test the generalizability of our findings. The period studied (1995-2004) represents a specific phase in the evolution of environmental reporting; subsequent developments in sustainability reporting standards and assurance practices may have altered the linguistic landscape of environmental disclosures. Additionally, while our computational approach enables analysis at scale, it necessarily simplifies some nuances of human interpretation that might be captured through qualitative discourse analysis.

Future research could extend our framework in several promising directions. Longitudinal analysis could track how linguistic strategies evolve in response to specific regulatory changes or environmental incidents. Cross-cultural comparison could examine how linguistic patterns vary across different institutional and linguistic contexts. Experimental studies could test how different disclosure formats affect stakeholder comprehension and decision-making. Finally, integration with quantitative environmental performance data could enable more direct analysis of the relationship between actual environmental outcomes and their linguistic representation.

In conclusion, this research demonstrates that artificial intelligence techniques can reveal systematic patterns in environmental communication that elude traditional content analysis.

By examining not just what companies say about their environmental performance but how they say it, we gain insight into the strategic dimensions of corporate environmental transparency. As environmental disclosure becomes increasingly important for corporate accountability and stakeholder decision-making, understanding these linguistic dimensions becomes essential for ensuring that disclosures serve their intended informative purpose rather than merely their presentational function.

References

Abrahamson, E., Amir, E. (1996). The information content of the president's letter to shareholders. *Journal of Business Finance Accounting*, 23(8), 1157-1182.

Cho, C. H., Patten, D. M. (2004). Environmental reporting and political economy: A theoretical framework. *Accounting, Auditing Accountability Journal*, 17(4), 492-515.

Clatworthy, M., Jones, M. J. (2003). Financial reporting of good news and bad news: Evidence from accounting narratives. *Accounting and Business Research*, 33(3), 171-185.

Davis, A. K., Tama-Sweet, I. (2002). Managers' use of language across alternative disclosure outlets: Earnings press releases versus MD&A. *Contemporary Accounting Research*, 29(3), 804-837.

Deegan, C., Gordon, B. (1996). A study of the environmental disclosure practices of Australian corporations. *Accounting and Business Research*, 26(3), 187-199.

Gray, R., Kouhy, R., Lavers, S. (1995). Corporate social and environmental reporting: A review of the literature and a longitudinal study of UK disclosure. *Accounting, Auditing Accountability Journal*, 8(2), 47-77.

Henry, E. (2004). Are investors influenced by how earnings press releases are written? *Journal of Business Communication*, 41(4), 349-367.

Li, F. (2004). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, 45(2-3), 221-247.

Merkl-Davies, D. M., Brennan, N. M. (2000). Discretionary disclosure strategies in corporate narratives: Incremental information or impression management? *Journal of Accounting Literature*, 26, 116-194.

Neu, D., Warsame, H., Pedwell, K. (1998). Managing public impressions: Environmental disclosures in annual reports. *Accounting, Organizations and Society*, 23(3), 265-282.