Faculty of Business and Economics
Department of Accountancy and Finance

Corporate narratives and textual analysis: Perspectives on top management's language use in financial and sustainability reporting

PhD thesis submitted for the degree of Doctor of Applied Economics at the University of Antwerp to be defended by Anil Berkin

Supervisor(s): Prof. Dr. Walter Aerts; Prof. Dr. Tom Van Caneghem

Antwerp, 2024

ISBN: 9789057288395
Deposit Number: D/2024/12.293/09
Printed by: Universitas Cursusdienst
FWO File Number: G071617N

Corporate narratives and textual analysis: Perspectives on top management's language use in financial and sustainability reporting

Ph.D. Dissertation
University of Antwerp
Faculty of Business and Economics
Department of Accountancy and Finance

Anil Berkin
Antwerp, 2024

Internal Jury Members

Prof. Dr. Ine Paeleman
Prof. Dr. Walter Aerts
Prof. Dr. Tom Van Caneghem
Prof. Dr. Ann Jorissen

External Jury Members

Prof. Dr. Raf Orens, KU Leuven
Prof. Dr. Maria Roszkowska-Menkes, Tilburg University
Prof. Dr. Reggy Hooghiemstra, University of Groningen

ACKNOWLEDGEMENTS

After a long period of my doctoral studies, I am grateful to have reached this significant milestone. Over the course of the past five years, I have encountered numerous enriching and enjoyable experiences, along with various challenges: office robbery, Covid-19 crisis and lockdowns, and witnessing the unfortunate event of the university building burning down. Luckily, survived all.

First, I extend my deepest gratitude to my supervisors: Prof. Dr. Walter Aerts and Prof. Dr. Tom Van Caneghem, for this invaluable opportunity to collaborate with them. Under their guidance, I have acquired a wealth of knowledge and insights, which I will carry all the time. I also thank my always positive and supportive chairperson: Prof. Dr. Ine Paeleman. I am sincerely appreciative of the guidance and mentorship provided by Prof. Dr. Maria Roszkowska-Menkes, who not only served as a jury member, or co-author but also as a friend. Their invaluable contributions have significantly enriched my Ph.D. experience. Furthermore, I extend my heartfelt gratitude to my esteemed jury members: Prof. Dr. Ann Jorissen, Prof. Dr. Raf Orens, and Prof. Dr. Reggy Hooghiemstra, who reviewed and provided constructive feedback for enhancing the quality of this dissertation. I would also like to thank Prof. Dr. Kris Hardies for his support and encouragement throughout this journey. I am grateful to Prof. Dr. David Martens for inspiring my interest in machine learning and AI technologies during my initial year at the University of Antwerp, shaping the foundation of the fourth chapter of this dissertation, and resulted in a journal submission. Furthermore, I wish to extend my appreciation to Prof. Dr. Karol Klimczak for his guidance and support during my short exchange at the Lodz University of Technology in Poland.

I am grateful for the amazing and special friendships over the course of my journey: Patrick Allmis, Dimitris Georgoudakis, Spiros Gkousis and Cathrin Schödler. Particularly for the time spent at the Ossenmarkt. Special gratitude is for my dear friends from the Faculty of Business and Economics: Tristan De Blick, Katja Bringmann, Pieter De Rijck, Jovana Cadenovic, Jef Van Cappellen, Michiel Dierckx, Katrien Jansen, Nina Marien, Jo Mentens, Kevin Van Mencxel, Tijmen van Kempen, Fynn Ohlrogge, Michiel Van Roy, Jonas Vandenbruaene, Jonas Vandennieuwenhuysen, Sahar Salehi and Ekin Yurdakul. A heartfelt thanks to Annelies Cosaert, whose path crossed with mine during the final stages of my Ph.D., for all the happy moments. I am grateful to my Belgian family: the Ceulemans and Beldé families. Especially my sister: Marte Beldé for her support since my arrival in Belgium in 2018.

My deepest gratitude goes to my family in Turkey: My mother Sevinc Erol and my father Koray Berkin; and my lovely grandparents. Without their support and encouragement, I would not be where I am today.

ABSTRACT

Ph.D. in Applied Economics: Corporate narratives and textual analysis: Perspectives on top management's language use in financial and sustainability reporting

The dissertation investigates top management's language use in financial and sustainability reporting, offering insightful perspectives and methodological innovations. It delves into differences in language use in distinct disclosure genres influenced by audience projection and communicative purpose, and the role of visibility in rhetorical impression management in sustainability reporting. By exploring these dimensions, the dissertation sheds light on how companies navigate communication challenges, adapt to regional level institutional settings and employ strategic approaches to shape perceptions and manage stakeholder relationships. Furthermore, the dissertation introduces feasibility of employing machine learning methods for attributional content and framing analysis in corporate reporting, highlighting the potential in enhancing narrative disclosure analysis. Collectively, these findings contribute to a deeper understanding of corporate communication practices and pave the way for future research endeavors in the field.

ABSTRACT IN HET NEDERLANDS

Doctor in de toegepaste economische wetenschappen: Bedrijfsverhalen en tekstanalyse: perspectieven op het taalgebruik van het topmanagement in financiële en duurzaamheids rapportage

Het proefschrift onderzoekt het taalgebruik van het topmanagement in financiële en duurzaamheidsrapportage en biedt inzichtelijke perspectieven en methodologische innovaties. Het proefschrift onderzoekt verschillen in taalgebruik in verschillende genres van rapportage, beïnvloed door publieksprojectie en communicatief doel, en de rol van zichtbaarheid in impressie management. Door deze dimensies te onderzoeken, werpt het proefschrift licht op hoe bedrijven omgaan met communicatie-uitdagingen, zich aanpassen aan institutionele instellingen op regionaal niveau en strategische benaderingen gebruiken om percepties vorm te geven en relaties met belanghebbenden te beheren. Bovendien introduceert het proefschrift de haalbaarheid van het gebruik van machine learning-methoden voor attributieve inhoud en framing-analyse in bedrijfsrapportage, waarbij het potentieel wordt benadrukt om de analyse van narratieve openbaarmakingsanalyse te verbeteren. Gezamenlijk dragen deze bevindingen bij aan een dieper begrip van de bedrijfscommunicatiepraktijken en banen ze de weg voor toekomstige onderzoeksinspanningen op dit gebied.

TABLE OF CONTENTS

CHAPTER I **10**

INTRODUCTION **10**

1. AIM OF THE DISSERTATION 10
2. CORPORATE NARRATIVES AND IM 10
3. LEGITIMACY THEORY 13
4. THE STRUCTURE OF THE DISSERTATION 15

CHAPTER II **19**

STUDY 1: LANGUAGE STYLE IN CORPORATE NARRATIVES: HOW REPORTING GENRE AND INSTITUTIONAL SETTING AFFECT LINGUISTIC STYLE IN CEO LETTERS **19**

1. INTRODUCTION 20
2. CEO LETTERS AS CORPORATE REPORTING GENRE AND LANGUAGE STYLE 23
   2.1. Regulatory context, audience diversity and evaluative legitimacy criteria 23
   2.2. The effect of region-based institutional setting on the informal and conversational storytelling style in SR CEO letters 29
3. DATA AND METHODOLOGY 33
   3.1. Sample selection and data collection 33
   3.2. Methods 36
4. RESULTS 44
   4.1. Descriptive statistics 44
   4.2. Multivariate regression analyses 48
   4.3. Additional analyses 56
5. DISCUSSION AND CONCLUSION 70

CHAPTER III **85**

STUDY 2: RHETORICAL IMPRESSION MANAGEMENT IN SUSTAINABILITY REPORTING CEO LETTERS: THE EFFECT OF GENERAL AND ISSUE VISIBILITY **85**

1. INTRODUCTION 86
2. LITERATURE REVIEW 90
   2.1. SR FROM A SOCIO-POLITICAL LENS 90
   2.2. Impression Management 92
3. HYPOTHESES DEVELOPMENT 94
   3.1. General visibility driven by sensitive industry membership and IM 95
   3.2. Issue visibility driven by sustainability controversies and IM 97
4. METHODOLOGY 100
   4.1. Sample selection and data collection 100
   4.2. Linguistic analysis 104
   4.3. Factor analysis 106
   4.4. Empirical models 111
5. RESULTS 114
   5.1. Descriptive statistics 114
   5.2. Multivariate analysis 118
   5.3. Robustness and extra analyses 122
6. DISCUSSION AND CONCLUSION 129

APPENDIX 1 142

CHAPTER IV **149**

STUDY 3: FEASIBILITY ANALYSIS OF MACHINE LEARNING FOR PERFORMANCE-RELATED ATTRIBUTIONAL STATEMENTS **149**

1. INTRODUCTION 150
2. LITERATURE REVIEW 155
   2.1. Management commentary and narrative disclosure research 155

2.2. Machine learning methods and their use in narrative disclosure research ... *162*
3. Methodology and Data ... 166
3.1. Training dataset ... *168*
3.2. Data preparation and text pre-processing ... *176*
3.3. Algorithm construction ... *178*
4. Results ... 180
4.1. Performance evaluation of the classifiers ... *180*
4.2. Supplementary analyses ... *191*
5. Discussion and conclusion ... 195

CHAPTER V ... 208

CONCLUSION ... 208

1. Theoretical Contributions ... 208
2. Practical Contributions ... 210
3. Limitations and future recommendations ... 212

CHAPTER I

INTRODUCTION

1. Aim of the Dissertation

The aim of this dissertation is to investigate the use of top management's language use in financial reporting (FR) and sustainability reporting (SR). The dissertation is situated at the intersection of accounting, linguistics and sustainability research. The investigation of this dissertation involves: (1) identification of linguistic style features and linguistic style in FR and SR narratives and conducting a comparative analysis; (2) examination of the influence of regional-level institutional settings, industry affiliation and ESG controversies on the utilization of impression management (IM) in SR; and (3) assessment of the viability of employing novel methodologies (i.e., machine learning) for IM research. In this dissertation, legitimacy theory serves as the primary framework to comprehend and elucidate the motivations behind the use of different linguistic styles and IM strategies in corporate narratives. Subsequent sections in this chapter provide the definitions of IM and legitimacy theory; providing a brief theoretical background employed in this dissertation; and outline the structure of the dissertation.

2. Corporate narratives and IM

Organizations utilize corporate disclosures as communication devices to convey information about organizational performance to external stakeholders, serving both informative and legitimacy-related purposes. Primary genres of corporate disclosure encompass financial reporting (FR), oriented towards presenting and contextualizing financial accomplishments and performance with a focus on economic and financial indicators, such as growth, profitability and operational performance, and SR[1] which focuses on providing information regarding a company's sustainability performance (SP) and commitment to social, environmental and ethical issues. The narrative sections of these disclosures are highly visible and discretionary in nature (Amernic & Craig, 2007). They establish the overall mood of the entire report and facilitate direct communication between top management and stakeholders (Caliskan et al., 2021). These sections offer valuable insights into company policy, strategy, commitment and accountability; and while they can be rational and appropriate, they can be also used to create favorable company images (Amernic & Craig, 2007). Consequently, they are particularly prone to IM.

Leary and Kowalski (1990, p. 34) define IM as *the process by which individuals attempt to control the impressions others form of them*, with the primary objective of being perceived as favourable as possible. In corporate reporting, seven different managerial IM strategies are identified: (1) reading ease manipulation; (2) rhetorical manipulation; (3) thematic manipulation; (4) visual and structural manipulation; (5) performance comparisons; (6) choice of earnings number; and (7) attribution of performance (Merkl-Davies & Brennan, 2007). Among these, rhetorical manipulation, which relates to linguistic style; thematic manipulation, encompassing selective thematic disclosure; and attribution of performance, concerning the utilization of causal statements are considered verbal IM, which can be utilized in narrative sections of corporate disclosures. In this dissertation, we use IM to refer to verbal IM. In the context of corporate disclosure, IM entails selectively presenting information in a manner designed to distort readers' perceptions of firm performance (Godfrey et al., 2003; Merkl-Davies & Brennan, 2007). Drawing from social psychology research, key motivators for IM

[1] In the literature, different names have been used for corporate social responsibility (CSR) reports; social/environmental reports; environmental, social and governance (ESG) reports; corporate social disclosures, corporate environmental reporting; social/environmental reporting etc. In this dissertation, we use "sustainability reporting", following the inclusive definition of sustainability in the Brundtland Commission's Report (1987).

include publicity, dependency on target and image gap as the main drives for IM (Leary & Kowalski, 1990). Higher levels of publicity (i.e., greater public visibility of one's behavior), increased dependency on the target audience (i.e., a desire for approval) and a larger image gap (i.e., discrepancy between the desired and current image) lead to heightened concerns about how one's behavior is perceived, thereby motivating individuals to exert control and influence over those perceptions (Leary & Kowalski, 1990).

There are various forms of IM depending on different motives. Tedeschi and Melburg's IM framework (1984) identifies subcategories of IM strategies as assertive and defensive, depending on different IM motives. Based on the framework, it is argued that companies use assertive tactics mainly to build legitimacy and defensive tactics address shortfalls in the company performance (Ogden & Clarke, 2005). Assertive IM is mainly conducted by entitlements (i.e., the attribution of positive outcomes to internal causes) and enhancements (i.e., highlighting positive outcomes in spite of negative external circumstances), while defensive mechanisms include excuses (i.e., the attribution of negative outcome to negative external factors), justifications (i.e., accepting the responsibility of a negative outcome but using this as a step to achieve higher goals) and causality denials (i.e., denying the responsibility of a negative outcome) (Zhang & Aerts, 2015). There are different subcategorizations of IM, as well. For example, Hooghiemstra (2000) used "acclaiming" and "accounting". According to Hooghiemstra (2000), acclaiming tactics include enhancements and entitlements, while accounting tactics include excuses and justifications. He further argued an additional layer, however, whether these tactics address responsibilities (as entitlements and excuses) or consequences (as enhancements and justifications). Bansal and Kistruck (2006) used "demonstrative" versus "illustrative" as IM labels, where the former relates to detailed explanations (e.g., with quantitative data or graphs) and the latter relates to less detailed information with higher use of images. Elsbach (1994) classified IM strategies as "accommodative" (i.e., acceptance of responsibility) versus "defensive" (i.e., insistence that a problem does not exist and actions to resume normal operations); and Higgins and Walker (2012) classified IM strategies as logos, ethos and pathos. This classification relates to the Aristotelean persuasion and communication appeals, where logos aims at convincing audiences by referring to "logic, data and evidence" using logical argumentation and facts to support a claim; ethos calls to the audiences' ethical norms, displays a more credible and trustable character and draws on expertise, deference and self-reference; and pathos focusing on provoking emotions (Brennan & Merkl-Davies, 2014; Higgins & Walker, 2012; Hossain et al., 2019)[2].

3. Legitimacy theory

We employ legitimacy theory to shed light on top management's language use in FR and SR. The theory posits that an organization's existence hinges upon its capacity to align with societal expectations, social norms, values and beliefs (Suchman, 1995, p. 574). Consequently, corporate actions should be perceived as desirable, proper and appropriate within social norms, values and beliefs (Suchman, 1995). Failure to comply with these social expectations may pose a threat to a company's survival (Deegan, 2002). One way to establish legitimacy is by shaping public perception through deliberate choices in presenting information to readers (i.e., IM) (Cho et al., 2015; Dunne et al., 2021; García-Sánchez & Araújo-Bernardo, 2019; Merkl-Davies & Brennan, 2007). Corporate narratives, such as CEO letters, serve as a means through which managers can control legitimacy and promote a trustworthy company image to some extent

[2] In the business ethics literature, IM also raises a number of important ethical questions, such as whether using defensive tactics is ethical or not (Provis, 2010). In this dissertation, it is emphasized that IM practices in corporate narratives are not supposed to be perceived as negative or unethical strategies, instead they serve both self-presentation and information-sharing purposes (Yan et al., 2019). Another question raised about IM in the literature is whether IM is implemented consciously or whether they are not aware of it. Hooghiemstra (2003) argues that in everyday behavior, people engage in IM unconsciously. At the corporation level, however, prior literature mainly shows they are strategic and tactical maneuvers (Hooghiemstra, 2003).

(Aerts & Cormier, 2009; De Villiers & Maroun, 2018; Fuoli, 2018). Doing that, managers tailor legitimacy strategies to specific target audiences with different objectives (Suchman, 1995). For example, maintaining already existing legitimacy or repairing the damaged legitimacy requires different IM strategies (Kuruppu et al., 2019).

Within the literature three main legitimacy types are identified: (1) pragmatic; (2) moral; and (3) cognitive (Suchman, 2005). In the context of organizational legitimacy, pragmatic legitimacy aims at organizations' immediate audiences, such as shareholders who are required to be persuaded through their self-interested need for financial returns (Edgar, 2018). Pragmatic legitimacy rests on the self-interested calculations leaving a large space for audiences to scrutinize organizational activities and their consequences (Suchman, 1995). Moral legitimacy focuses on societal stakeholders and requires managers to show that the company is doing the "*right thing*" from a moral perspective and aligning corporate performance outcomes with expected ethical practices (Marais, 2012). In corporate communication, humanism, benevolence, diversity and openness to others are good examples of this type of legitimacy (Marais, 2012). Depending on communication goals, the aim can be on different legitimacy types. For example, Marais (2012) posits that managerial focus predominantly centers on achieving moral legitimacy when communicating sustainability efforts. Finally, cognitive legitimacy concerns about the assumptions and awareness of audiences about companies. Cognitive legitimacy involves claims about the willingness to follow widely accepted norms and standards (Hossain et al., 2019). Unlike pragmatic and moral legitimacy, cognitive legitimacy does not depend on public discussions, where people can evaluate benefits and ethical aspects (Suchman, 1995). As cognitive legitimacy relies on unspoken assumptions (Suchman, 1995), this dissertation mainly focuses on pragmatic and moral legitimacy types.

4. The structure of the dissertation

In line with the aim of the dissertation, we first compare and contrast linguistic style features and rhetorical profiles in FR and SR narratives. We argue that the genre-specific features and different audience expectations result in different linguistic styles in FR and SR. Moreover, we argue that regional-level institutional setting also plays a role in different linguistic styles. For our comparison, we use quantitative methods, e.g., multivariate regression models, and a dictionary-based approach (i.e., pre-defined words and rules) using Linguistic Inquiry and Word Count – 22 (LIWC-22) software (Boyd et al., 2022). This study helps us explore the distinct style-related features in both genres and define rhetorical appeals apparent. In the next study, we examine the effect of visibility on the use of IM in SR. We argue that different visibility notions, i.e., general visibility driven by industry membership and issue visibility driven by controversies, result in different IM strategies. This allows us to understand what factors are behind different IM strategies, and how they are affected by different legitimacy purposes. Similar to the first study, we use multivariate regression models to test the effect of general and issue visibility on the use of IM. In addition to LIWC-22 analysis, this study also controls for thematic differences in corporate narratives using an unsupervised machine learning algorithm. While both studies predominantly rely on dictionary-based methodologies, we argue that this research methodology remains to enhance for a more comprehensive understanding of context, extending beyond mere word counts, through the incorporation of machine learning algorithms. In the final study, we test the feasibility of using machine learning algorithms on causal statements, that are complex in nature (El Haj et al., 2019; Berkin et al., 2023).

Theoretically, this dissertation emphasizes the relevance of legitimacy theory in understanding corporate communication behavior. It highlights the need to distinguish between different legitimacy and stakeholder motives. The results of this dissertation show distinct language use across FR and SR, and motives behind IM in SR, supporting the theory's relevance, due to different legitimacy concerns of targeted audiences, i.e., moral and pragmatic legitimacy. Specifically, the empirical findings of this dissertation offer insights to enhance the quality of SR guidelines, such as GRI or CSRD directives, narrowing information asymmetry in narrative sections. Our emphasis on transparency and accountability shows the need for regulatory bodies to update guidelines, considering regional and industry-specific factors. The insights of this dissertation are expected to aid stakeholders to critically evaluate narrative sections of corporate reports to distinguish between opportunistic storytelling and genuine sustainability commitment. Additionally, the dissertation shows that machine learning algorithms can improve the efficiency of text analysis, helping investors and auditors in comprehensively assessing corporate communication and identifying potential red flags. An in-depth examination of the referred studies will be conducted across subsequent chapters. The conclusive section then addresses the contributions, limitations and outlines potential avenues for future research.

Reference List

Aerts, W., & Cormier, D. (2009). Media legitimacy and corporate environmental communication. *Accounting, Organizations and Society, 34*(1), 1-27. https://doi.org/10.1016/j.aos.2008.02.005

Amernic, J. H., & Craig, R. J. (2007). Guidelines for CEO-speak: editing the language of corporate leadership. *Strategy & Leadership, 35*(3), 25-31. https://doi.org/10.1108/10878570710745802

Bansal, P., & Kistruck, G. (2006). Seeing Is (Not) Believing: Managing the Impressions of the Firm's Commitment to the Natural Environment. *Journal of Business Ethics, 67*(2), 165-180. https://doi.org/10.1007/s10551-006-9021-9

Berkin, A., Aerts, W., & Van Caneghem, T. (2023). Feasibility analysis of machine learning for performance-related attributional statements. *International Journal of Accounting Information Systems, 48*. https://doi.org/10.1016/j.accinf.2022.100597

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). *The Development and Psychmetric Propoerties of LIWC-22*. The University of Texas at Austin.

Brennan, N. M., & Merkl-Davies, D. M. (2014). Rhetoric and argument in social and environmental reporting: the Dirty Laundry case. *Accounting, Auditing & Accountability Journal, 27*(4), 602-633. https://doi.org/10.1108/aaaj-04-2013-1333.

Caliskan, A. O., Esen, E., & Barkemeyer, R. (2021). Impression management tactics in the CEO statements of Turkish sustainability reports. *Business Ethics, the Environment & Responsibility, 30*(4), 485-506. https://doi.org/10.1111/beer.12374

Cho, C. H., Michelon, G., Patten, D. M., & Roberts, R. W. (2015). CSR disclosure: the more things change…? *Accounting, Auditing & Accountability Journal, 28*(1), 14-35. https://doi.org/10.1108/aaaj-12-2013-1549

De Villiers, C., & Maroun, W. (2018). *Sustainability accounting and integrated reporting*. Routledge.

Deegan, C. (2002). The legitimising effect of social and environmental disclosures – a theoretical foundation. *Accounting, Auditing & Accountability Journal, 15*(3), 282-311. https://doi.org/10.1108/09513570210435852

Directive (EU) 2022/2464 of the European Parliament and of the Council of 14 December 2022 amending Regulation (EU) No 537/2014, Directive 2004/109/EC, Directive 2006/43/EC and Directive 2013/34/EU, as regards corporate sustainability reporting. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022L2464.

Dunne, N. J., Brennan, N. M., & Kirwan, C. E. (2021). Impression management and Big Four auditors: Scrutiny at a public inquiry. *Accounting, Organizations and Society, 88*. https://doi.org/10.1016/j.aos.2020.101170

Edgar, V. C., Beck, M., & Brennan, N. M. (2018). Impression management in annual report narratives: the case of the UK private finance initiative. *Accounting, Auditing & Accountability Journal, 31*(6), 1566-1592. https://doi.org/10.1108/aaaj-10-2016-2733

El-Haj, M., Rayson, P., Walker, M., Young, S., & Simaki, V. (2019a). In search of meaning: Lessons, resources and next steps for computational analysis of financial discourse. *Journal of Business Finance & Accounting, 46*(3-4), 265-306. doi:10.1111/jbfa.12378

Elsbach, K., D. (1994). Managing Organizational Legitimacy in the California Cattle Industry: The Construction and Effectiveness of Verbal Accounts. *Administrative Science Quarterly, 39*(1), 57-88. https://doi.org/10.2307/2393494

Fuoli, M. (2018). Building a Trustworthy Corporate Identity: A Corpus-Based Analysis of Stance in Annual and Corporate Social Responsibility Reports. *Applied Linguistics, 39*(6), 846-885. https://doi.org/10.1093/applin/amw058

García-Sánchez, I. M., & Araújo-Bernardo, C. A. (2019). What colour is the corporate social responsibility report? Structural visual rhetoric, impression management strategies, and stakeholder engagement. *Corporate Social Responsibility and Environmental Management, 27*(2), 1117-1142. https://doi.org/10.1002/csr.1869

Godfrey, J., Mather, P., & Ramsay, A. (2003). Earnings and Impression Management in Financial Reports: The Case of CEO Changes. *Abacus, 39*(1), 95-123. https://doi.org/10.1111/1467-6281.00122

Higgins C. & Walker, R. (2012). Ethos, logos, pathos: Strategies of persuasion in social/environmental reports. *Accounting Forum, 36*(3), 194-208. https://doi.org/10.1016/doi.org/j.accfor.2012.02.003

Hooghiemstra, R. (2000). Corporate Communication and Impression Management – New Perspectives Why Companies Engage in Corporate Social Reporting. *Journal of Business Ethics, 27*, 55-68. https://doi.org/10.1023/A:1006400707757

Hooghiemstra, R. (2003). *The construction of reality: cultural differences in self-serving behaviour in accounting narratives*. Paper presented at the ERIM PhD Series Research in Management, Rotterdam.

Hossain, M., Islam, M. T., Momin, M. A., Nahar, S., & Alam, M. S. (2019). Understanding Communication of Sustainability Reporting: Application of Symbolic Convergence Theory (SCT). *Journal of Business Ethics, 160*(2), 563-586. https://doi.org/10.1007/s10551-018-3874-6

Kuruppu, S. C., Milne, M. J., & Tilt, C. A. (2019). Gaining, maintaining and repairing organisational legitimacy. *Accounting, Auditing & Accountability Journal, 32*(7), 2062-2087. https://doi.org/10.1108/aaaj-03-2013-1282

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin, 107*(1), 34–47. https://doi.org/10.1037/0033-2909.107.1.34
Marais, M. (2012). CEO rhetorical strategies for corporate social responsibility (CSR). *Society and Business Review, 7*(3), 223-243. https://doi.org/10.1108/17465681211271314
Merkl-Davies, D. M., & Brennan, N. M. (2007). Discretionary disclosure strategies in corporate narratives: incremental information or impression management? *Journal of Accounting Literature, 26*(116-196).
Ogden, S., & Clarke, J. (2005). Customer disclosures, impression management and the construction of legitimacy. *Accounting, Auditing & Accountability Journal, 18*(3), 313-345. doi:10.1108/09513570510600729
Provis, C. (2010). The ethics of impression management. *Business Ethics: A European Review, 19*(2), 199-212. https://doi.org/10.1111/j.1467-8608.2010.01584.x
Suchman, M. C. (1995). MANAGING LEGITIMACY: STRATEGIC AND INSTITUTIONAL APPROACHES. *Academy o/ Management Review, 20*(3), 571-610. https://doi.org/10.5465/amr.1995.9508080331
Tedeschi, J. T., & Melburg, V. (1984). Impression management and influence in the organization. In S. B. Bacharach, & E. J. Lawler (Eds.), *Research in the sociology of organizations* (pp. 31-58). JAI Press.
Yan, B., Aerts, W., & Thewissen, J. (2019). The informativeness of impression management − financial analysts and rhetorical style of CEO letters. *Pacific Accounting Review, 31*(3), 462-496. https://doi.org/10.1108/par-09-2017-0063
Zhang, S., & Aerts, W. (2015). Management's performance justification and failure to meet earnings thresholds. *Accounting and Business Research, 45*(6-7), 841-868. https://doi.org/10.1080/00014788.2015.1048771

CHAPTER II

STUDY 1: LANGUAGE STYLE IN CORPORATE NARRATIVES: HOW REPORTING GENRE AND INSTITUTIONAL SETTING AFFECT LINGUISTIC STYLE IN CEO LETTERS

Abstract

We examine the language used by top management in CEO letters incorporated in FR and SR as distinct genres. Differences in regulatory context, projected audience and evaluative legitimacy criteria ground distinct shared conventions and expectations, affecting a CEO's linguistic style in each reporting genre. We expect those shared conventions and expectations to be anchored more in an accountability logic and related formal reasoning in FR CEO letters and more storytelling and affective positioning in SR CEO letters. Moreover, we expect that regional-level institutional setting to affect linguistic style given differences in implicit versus explicit CSR cultures. Focusing on patterns of linguistic indicators, we identify and reveal distinct linguistic styles in FR and SR CEO letters, significantly calibrated by institutional environment.

1. Introduction

Companies use corporate disclosures as communication devices to signal organizational performance to external constituents for informative and legitimacy purposes. Such corporate disclosures may originate from different settings with distinct shared conventions and expectations. Primary disclosure genres are FR, which is a corporate disclosure genre aiming at presenting and contextualizing financial achievements with a focus on corporate governance and financial metrics, such as growth, profitability and operational performance, and SR which focuses on providing information regarding a company's SP and commitment to social, environmental and ethical issues (Băndoi et al., 2021; Barkemeyer et al., 2014; Mäkelä & Laine, 2011). Disclosure genres, such as FR and SR, tend to facilitate communication between writers and their audience through shared conventions and expectations about what to say and how to say it in recurring rhetorical situations (Fuoli, 2018). Regardless of differences in shared conventions and expectations, language choices in both FR and SR narratives remain to a large extent discretionary, holding potential to purposefully build, maintain or change corporate image and reputation in their respective field of interest, known also as IM (Brennan & Merkl-Davies, 2014; Martins et al., 2021; Merkl-Davies et al., 2011).

CEO letters in FR and SR stand out as subgenres that are highly visible narratives and are particularly prone to IM due to their voluntary and discretionary nature (Barkemeyer et al., 2014; Caliskan et al., 2021; Fuoli, 2018). They are common in both disclosure genres as introductory sections of the main narratives, set the tone for the entire report and establish direct communication between top management and readers (Caliskan et al., 2021; Hyland, 1998; Mäkelä & Laine, 2011; Skulstad, 2005). They provide personalized messages and business tales of top management regarding the most important and significant corporate events, achievements and future developments as claimed by the corporate leader (Amernic & Craig, 2007). They can be rational and appropriate, provide convincing arguments or create favorable company images (Amernic & Craig, 2007). They hold significant symbolic value in bringing sense to the way the company develops, in structuring authority and leadership and in building stakeholder support (Barkemeyer et al., 2014; Caliskan et al., 2021; Leibbrand, 2015). Prior research investigating CEO letters as a *subgenre* in corporate disclosures highlights significant differences that relate to the communication goals between writers and readers. For example, Hyland (1998) distinguishes the CEO letter and the directors' report of the same FR as different disclosure subgenres and shows that CEOs prefer to display more alignment with their readers and inform them about their intentions and attitudes, while in the directors' report, there is lower need to intervene in the discourse. Similarly, Garzone (2005) demonstrates a more personalized use of language in CEO letters compared to the rest of FR. She argues that CEO letters provide writers with a storytelling space to express their stance in an evaluative way, potentially relying on a broad repertoire of IM tactics in order to personally connect and align with the audience. Given this personalized communicative setting, FR and SR CEO letters are considered separate subgenres.

The language used in corporate narratives, particularly CEO letters in FR and SR, has received growing attention from both business and linguistic researchers. In CEO letters, top management is expected to support message trustworthiness and establish credibility among readers by using language that is congruent with both communicative intent and audience expectations which may be significantly different for FR versus S (Jonäll & Rimmel, 2010). Analyzing and contrasting language between FR and SR may offer valuable insights into how corporate leaders utilize these reports to calibrate implied worldviews (Mäkelä & Laine, 2011), and IM tendencies (Caliskan et al., 2021) according to reporting genre. While some studies (e.g., Barkemeyer et al., 2014; Blanc et al., 2019; Caliskan et al., 2021; Fuoli, 2018; Mäkelä & Laine, 2011) shed light on aspects of language use in FR and SR CEO letters, a systematic understanding of how differences in implicit conventions and expectations regarding topical content, information-sharing motives, target audience characteristics and institutional setting affect language in FR versus SR, is still lacking. In addition, studies investigating language in FR and SR CEO letters are mostly characterized by small sample size (e.g., case study), a limited number of linguistic features (e.g., emotion markers), a focus on one industry (e.g., mining) or country (e.g., US) and mainly embedded in a qualitative methodology (e.g., Barkemeyer et al., 2014; Blanc et al., 2019; Caliskan et al., 2021; Fuoli, 2008; Mäkelä & Laine, 2011). Our research overcomes these limitations by employing a relatively broad set of linguistic markers and a quantitative approach to textual data from various industries and countries. Our investigation is guided by the following empirical questions: (1) What are the dominant linguistic style features in FR and SR CEO letters?; (2) How and why does the linguistic style in FR and SR narratives differ?; and (3) How does region-based institutional setting affect the use of language in respective narrative disclosures?

Theoretically, our research relies on legitimacy theory and institutional theory, which helps us explore how respective implied conventions and expectations of CEO letter genre depend on differences in regulatory context (including target audiences and disclosure expectations, evaluative legitimacy concerns and region-based institutional environment. These institutional variables set the broader framework within which top management exerts discretion in style-based IM. Empirically, we use a matched dataset of CEO letters in FR and SR, which allows us to compare linguistic style features in FR and SR narratives generated by the same management team for the same period. Our dataset consists of a total of 2,472 CEO letters (1,236 from stand-alone SR and 1,236 from annual FR) of 291 different companies from the EU, the UK, and the USA for the period 2010 up to 2019. Our corpus includes approximately 4 million words. We employ factor analysis on the selected linguistic variables (based on prior literature) to proxy style-based IM tendencies and identify a rational-factual language, a personalized sense-giving language and an assertive relationship-oriented language style as dominant linguistic profiles.

Our investigation shows how accountability and transparency are actively practiced through these two distinct subgenres. In addition, the interaction between subgenre and regional-level institutional settings offers further insights into how these settings may influence opportunistic storytelling behavior. This synthesis facilitates an enriched comprehension of language usage and communication strategies within distinct subgenres. Consequently, our work extends practical insights that hold relevance for analysts, investors and policymakers. Finally, by identifying FR and SR CEO letters as distinct subgenres with their respective promotional functions, our comparative approach allows us to both validate and contrast the most significant linguistic techniques employed in crafting FR and SR discourse. These findings hold significance for future research, providing a foundation for comprehensive exploration about the linguistic differences between these subgenres.

The remainder of the paper is organized as follows: Section 2 presents the theoretical framework leading to our research hypotheses. Section 3 introduces our data collection and methodology. Results and additional statistical analyses are presented in Section 4. Finally, in Section 5, we discuss our findings, present our contributions and limitations and provide future recommendations.

2. CEO letters as corporate reporting genre and language style

2.1. Regulatory context, audience diversity and evaluative legitimacy criteria

Organizations use corporate disclosures as communicative instruments aimed at signaling their operational performance to external constituents, serving both informational and legitimacy objectives. These corporate disclosures emanate from various contexts characterized by distinct expectations of audiences regarding what information to convey and how to convey it (Hyland, 2005, p. 88). FR and SR stand out as the two major genres of corporate reporting (Fuoli, 2018). Narrative sections of these corporate disclosures, such as CEO letters, retain a considerable degree of discretion, offering the opportunity to shape corporate image and reputation (Barkemeyer et al., 2014; Caliskan et al., 2021; Fuoli, 2018).

A key difference between FR and SR CEO letters is the regulatory context from which they emanate. Core characteristics of regulatory context on which they differ, relate to disclosure repertoire and target audience (Băndoi et al., 2021; Barkemeyer et al., 2014; Mäkelä & Laine, 2011). These differences are expected to affect the type of legitimacy concerns that CEOs would be concerned with in their narratives and the type of language that CEOs would use to address those concerns. FR CEO letters are an offshoot from a formal and well-regulated FR framework with a proclaimed logic of stewardship and accountability. FR CEO letters are expected to be somewhat aligned with the FR package that they accompany and, thus, be affected by an accountability logic with its prominent descriptive and logical reasoning (Yan & Aerts, 2014). The confines of corporate financial performance, with its correlates of profitability, financial health and growth, are well-established and through detailed generally accepted accounting rules which govern how corporate events are recognised, measured and disclosed. The FR framework is established on the premise of financial stakeholders (shareholders and creditors) as the primary audience. In this regard, FR CEO letters may also be expected to aim primarily at readers who tend to prioritise issues of financial competence and corporate governance, such as shareholders, investors and financial analysts (Băndoi et al., 2021; Barkemeyer et al., 2014; Fuoli, 2018; Leibbrand, 2015). This type of audience typically has the technical expertise to interpret formal governance and financial accounting language and they demand accountability for their own good (Cormier et al., 2009). A primary audience like this would grant legitimacy only if their particular interests are served by the company (Hossain et al., 2019).

In a FR context, the relationship between the CEO and the target audience is highly pragmatic, leaning towards the concept of pragmatic legitimacy, with its preoccupations of effectiveness, efficiency and risk/return relationships (Hossain et al., 2019; Marais, 2012). According to Suchman (1995, p.578), "*pragmatic legitimacy rests on the self-interested calculations of an organization's most immediate audiences*". It involves direct interdependencies between an organization and its audience, leading to higher scrutiny from the audience and more definite expectations of focused disclosure and communication (Suchman, 1995). A CEO would be expected, as a good financial steward, to seek pragmatic legitimacy by presenting an image perceived as honest and trustworthy but well-aligned with the pragmatic values of its primary financial audience and embedded in formal language that is taken for granted in such a context. In that vein, CEO language in FR CEO letters is expected to resort to financial accounting and governance language to a significant degree and present achievements and courses of action with supporting data and logical evidence, making FR CEO letter's expected messages fairly descriptive and analytic/rational in nature.

The SR regulatory context, on the other hand, is characterized by the relative lack of generally accepted standards for SR (in comparison with FR) and by a poorly institutionalized tradition of how SR should regulate stakeholder relationships. Two key issues that differentiate a SR context from a FR context are the lack of well-defined pragmatic performance criteria and audience diversity. A CEO is expected to respond to a societal sustainability call and cope with a range of stakeholders from a broad angle (e.g., customers, labour, civil society, NGOs) with their divergent needs and expectations (Băndoi et al., 2021; Barkemeyer et al., 2014; Fuoli, 2018; Lindgren et al., 2021; Mäkelä & Laine, 2011). In practice, SR CEO letters tend to elaborate on issues like the company's vision of its relationship with the environment and society at large, company-specific sustainability strategies or initiatives, as well as integration of sustainability into business strategies and models (Boiral et al., 2020), and this to different degrees of depth at the CEO's discretion.

The relative lack of generally accepted standards of SR is partly caused by the equivocal nature of sustainability performance. The multifaceted and somewhat opaque nature of the sustainability concept makes the measurement and comparison of sustainability indicators and the disclosure of reliable information challenging (Boiral et al., 2020) Some sustainability performance dimensions are highly qualitative in nature and almost impossible to quantify (Caliskan et al., 2021). The equivocality of the sustainability concept and the relative inability to pin down what sustainability performance really means also hinders the institutionalization of taken for granted control mechanisms and monitoring tools.

Audience diversity without implied hierarchy is also not helpful in establishing a generally accepted set of pragmatic performance standards around which expected disclosure templates could crystallize. Reconciling the interests of customers, labor organizations, suppliers and the community at large is never evident. The diversity of target audiences also means different levels of background knowledge, education and expertise on sustainability-related issues, which may cause management to use a more generic conversational style with simpler language and less reliance on detailed information and quantification in order not to burden the comprehensibility of the message for the average reader (Järvinen et al., 2020; Puroila & Mäkelä, 2019). Indices, such as CO2 emissions, can be difficult to interpret and compare by non-expert stakeholders, as they are still largely unstandardized (Boiral et al., 2020).

In such a context, management may be expected to shift focus from pragmatic legitimacy criteria to moral legitimacy concerns, as "*moral legitimacy rests not on judgements about whether a given activity benefits evaluator, but rather on judgements about whether the activity is the right thing to do*" (Suchman, 1995, p.579). As Suchman (1995) argues, moral legitimacy is rooted in beliefs about whether an activity aligns with societal welfare as defined by stakeholder values. When there are multiple and diverse stakeholders, managers may try, for moral legitimacy purposes, to connect with stakeholders on a more personal and emotional level, by utilizing storytelling techniques in SR CEO letters (Marais, 2012). Presenting the company, through a discourse of goodwill, as a good corporate citizen and a credible environmental steward is not uncommon in this respect (Bhatia, 2012; Lin-Hi & Blumberg, 2016). Emotional arguments, akin to discourse of 'values rhetoric', encompass thematic elements such as humanism, benevolence, diversity and openness to others, which seek to establish moral legitimacy through invoking emotions and affection to seduce the audience (Marais, 2012). Similarly, Araujo and Kollat (2018) argue that communication regarding sustainability-related information is permeated with emotions, morals and affective states, which bring stakeholders to subjectively experience the values of corporations through storytelling. Such personal and emotional discourse may help managers to humanize the information presented, making it relatable and conversational with diverse stakeholder groups. By telling stories of organizational achievements and their relationship with social values in a positive light, credible arguments are offered to legitimize organizational activities (Martins et al., 2021).

Relatedly, several studies provide evidence that SR is often based on highly readable qualitative language and abstract statements (Du & Yu, 2020; Hummel & Rötzel, 2019), whereby readers are guided to a preferred framing and interpretation of events and achievements. Consistent with this, Skulstad (2005) argues and shows that there is more need for directive and assistive language in SR introductions compared to FR introductions. She posits that FR readers (such as analysts and shareholders as professional users) have specific and clearer content expectations and are familiar with the institutional setting and regulatory aims of FR documents, while SR readers are generally more diverse, less professional in nature and require more guidance and directive language[3]. Moreover, focusing on audience differences, Fuoli (2018) compares discourse in FR and SR to explore how managers build trust with readers. His results show that language in FR is more rational, cautious and competent, while the language used in SR tends to be more committed, honest and caring. In addition, he shows that SR tends to reflect more affective status which may help foster a connection with readers, and modals and verbs may help to project integrity, as well as ethical commitment. Using these markers, managers can effectively tell a story about the company's ethical and responsible journey. Overall, Fuoli (2018) argues that these differences relate to varying reader profiles and their expectations. Similarly, Blanc et al. (2019) show that managers tend to adopt rational façade in FR and reputational façade in SR, mainly to cater to the tastes and needs of shareholders in FR, and to accommodate a much broader group of stakeholders in SR, where legitimacy concerns are divergent.

On many occasions, prior SR research made the case that SR is mainly used for reputation management purposes through greenwashing or rhetorical IM, rather than as a tool for environmental or societal accountability (Cho et al., 2012; Cho & Patten, 2007). Prior research has documented highly discretionary SR practices, leaving top management more degrees of freedom and a subjective space to exploit SR CEO letters for opportunistic storytelling and acclaiming behavior. Selective reporting and qualitative language can be used to neutralize or neglect potential detrimental effects of business activities on biodiversity or

[3] Skulstad (2005) investigates the use of action markers (e.g., "*we highlight*", "*we report*") and previews (e.g., "*as following*", "*we show below*") as textual markers within subgenres, which are found to be particularly different. Action markers in SR are used to indicate that the writer will be explaining and emphasizing certain issues later in the text, while in FR, they are used to indicate the discourse of act such as reporting (e.g., "*we report*", "*regret to/pleased to report*"). The second textual practice that differs in subgenres are previews, such as "*we show below*", "*as the following pages reveal*", that provide readers with clear explanations of the purposes behind publishing these reports, making it easier for readers to understand (Skulstad, 2005). In SR, they help facilitating reader's processing of the report, while in FR, they tend to be used to state the main knowledge and make claims, establishing directly manager's position (Skulstad, 2005). In SR, the way how action markers and previews are expected to enhance storytelling by providing readers with a roadmap of what to expect in the narrative, indicating the direction the narrative will take and help the diverse audience of SR process and engage with the narrative more effectively.

conceal poor sustainability performance. Assertive claims through expressions of sustainability commitment and selective reporting on sustainability initiatives are much more prevalent in SR CEO letters than actually commenting on sustainability performance metrics (Boiral, 2016; Hahn & Lülfs, 2014; Stacchezzini et al., 2016; Talbot & Boiral, 2018). Although self-serving management tendencies could be equally important for FR and SR CEO letters, the nature of IM tactics is expected to be significantly different, with higher propensity for propositional IM (e.g., attributional IM and selectivity in performance comparisons) in FR CEO letters and more style-based IM in SR CEO letters.

In line with these arguments, we expect language style in FR CEO letters to be more aligned with the preoccupations of highly regulated FR requirements with their logic of stewardship and accountability, which would promote a rational rhetorical appeal with more formal, rational-descriptive framing of language. For SR CEO letters, we expect the room for discretion in presenting an idiosyncratic story to be much more extensive, considering the relative paucity of regulatory requirements, the equivocality of the subject matter and the diversity of the audience. In SR CEO letters, the rhetorical need to align with and inform a broader audience repository is expected to promote a more informal, conversational storytelling style. Hence:

H1a: *In SR CEO letters, top management tends to use a more informal and conversational storytelling language style compared to FR CEO letters.*

H1b: *In FR CEO letters, top management tends to use a more formal, descriptive and rational language style compared to SR CEO letters.*

2.2. The effect of region-based institutional setting on the informal and conversational storytelling style in SR CEO letters

Scholars have long emphasized the influence exerted by country- and region-specific institutional environments on corporate governance and disclosure regulation and how this transpires through disclosure practices (Jackson & Apostolakou, 2010; Rim et al., 2024; Yan & Aerts, 2014). Hence, we posit that region-specific institutional setting may play a pivotal role in how top management crafts storytelling in SR CEO letters.

Prior comparative studies on SR have found substantial differences in reporting practices across regions. For example, Ferri (2017), comparing SR from Italy, Brazil and the USA, shows that SR content is influenced by the characteristics of the national institutional context along political, cultural, religious and legal lines. Fifka and Drabble (2012) demonstrate that the cultural and socio-economic environment in Finland and the UK has an impact on the extent and focus of disclosure. They argue that countries, where sustainability is strongly embedded in legislation, allow less space for voluntary reporting choices than in liberal market economies with lower regulation. Rosati and Faria (2019) provide empirical evidence that various institutional factors, such as political and legal system dimensions, as well as socio-cultural traits (e.g., individualism, power distance) significantly affect how companies report on sustainable development goals. More recently, Rim et al. (2024) provide empirical evidence on the link between CSR transparency strategies, perceived altruism and skepticism, and corporate trust and how these differ among country-based cultures.

Following, Matten and Moon (2008) argue that companies tend to adapt their CSR practices to the expectations and institutional background of the region they operate in. They observe, for example, that US firms proactively claim and communicate their CSR practices significantly more than European firms. They proposed the concept of "*implicit*" versus "*explicit*" CSR in order to compare and contrast how companies perceive and implement their business responsibility to society. According to the framework, firms' CSR practices will vary along an implicit-explicit continuum reflecting national social expectations as incorporated in national institutional disparities (LaGore et al., 2020; Matten & Moon, 2008). The implicit versus explicit CSR perspective has significant implications for CSR communication and both intent behind communication and language content and style would be affected (Lee & Riffe, 2019; Matten & Moon, 2008).

A key argument of Matten and Moon's (2008) framework is that national business systems ground the implicit versus explicit nature of CSR because the government, corporations, and markets that define the underlying institutional framework set the norms, incentives, and rules of CSR that companies respond to. From that perspective, they argue that the US-style CSR is explicit. The US-style CSR is grounded in a system that provides incentives and opportunities for companies to combine social and societal interests with business value through voluntary programs and strategies to address issues considered to be the social responsibility of the company (Blindheim, 2012). It allows much more room for corporate initiative and tends to translate CSR issues into corporate policies and strategies while giving priority to firm-level identity and reputation as a starting point to address perceived expectations of specific stakeholder groups of the company (Blindheim, 2012).

In Europe, the CSR style is rather implicit where European companies are less likely to set up their own CSR agendas (Matten & Moon, 2008). The institutional framework of norms, values, and rules that they respond to are the result of "*coordinated approaches to economic and social governance*" through largely government–led partnerships (Matten & Moon 2008, p.410; Blindheim, 2012). They tend to be motivated by a need for a societal consensus on the legitimate roles and obligations of different groups of stakeholders. Many of the elements of implicit CSR occur in the form of codified norms, rules, and laws[4]. As these societal norms, networks, organizations, and rules result in requirements for companies, they generate rather implicit implications for the social responsibilities of business. Implicit CSR consists of the values and norms that define the obligations of corporate actors in "*collective rather than*

[4] For example, the EU directives (such as "2014/95/EU") provide additional governance on the consistency and comparability of non-financial information disclosed for firms located in the EU (Pizzi et al., 2021).

individual terms" (Matten & Moon 2008, p.409). In such a context, companies are confronted with considerable preset checks and balances, leaving much less room for companies to articulate their own social responsibilities (Blindheim, 2012).

The intent of explicit CSR is different in that it is deliberate, voluntary, and often strategic in nature, with significant incentives for self-serving IM in order to establish and assure firm-level CSR identity and reputation. On the other hand, the intent of implicit CSR would not really be reflective of a corporate decision, but would rather be a reaction to, or reflection of the company's institutional environment such that codified norms, rules, and laws reflect broader societal interests. While both types of CSR may have codified requirements, implicit CSR highlights societal norms, networks, institutions, and rules rather than the firm-driven practices and policies of explicit CSR. As such, explicit CSR would promote communicating specific company-level policies and practices to stakeholders and implicit CSR would put more emphasis on the obligations of corporate actors in "*collective rather than individual terms*" (Matten & Moon 2008, p.409), stressing relationship-building and consensus searching.

Overall, a US-style CSR would rely more on firm-level discretion in building an idiosyncratic CSR identity, a strong incentive for self-serving IM. As prior research suggests, the absence of implicit checks and balances and a tendency to claim and emphasize company-level policies and initiatives rather than reactively responding to what is institutionally expected would generally promote assertive acclaiming and positive tone management. This brings us to propose the following hypotheses:

H2a: *In US SR CEO letters, companies tend to be less focused on consensual relationship-building than in EU SR CEO letters.*

H2b: *In US SR CEO letters, companies tend to use more assertive storytelling than in EU SR CEO letters.*

The next section describes the methodology and the variables we investigate.

3. Data and methodology

3.1. Sample selection and data collection

We use a dataset of matched CEO letters in FR and SR (i.e., CEO letter FR and SR belonging to the same firm-year observation), which allows us to compare textual properties in FR and SR CEO letters prepared by the same management for the same period. To test our hypotheses, we initially gathered data from the EU, UK and USA listed companies that published a stand-alone SR for the fiscal year 2016 according to the Thomson Reuters' Asset4ESG (Refinitiv/Eikon) database. Subsequently, we extended this dataset for the same companies, encompassing the period from 2010 up to 2019 and manually obtain the reports in PDF from firm websites, the corporate register database (corporateregister.com), and the GRI global reporting database (globalreporting.org). We exclude companies operating in healthcare, financials, real estate sectors; companies operating in governmental and administrative activities, and academic and educational industries; companies that do not publish their reports in English; companies that follow integrated reporting; reports that do not include CEO letters; reports that are protected and cannot be processed; and CEO letters counting less than 350 words[5]. Following prior research (Clarkson et al., 2008; Cong et al., 2014; Fehre & Weber, 2016; Na et al., 2020), we manually processed CEO letters in SR, and excluded the ones that do not include sustainability information. This selection process provides us with a total of 2,472 CEO letters (1,236 from FR and 1,236 from stand-alone SR for the corresponding firm-year observation) from 291 different companies, consisting of approximately 4 million words.

[5]The Receptiviti API generates scores on language using different measures, such as personality, emotions or cognition. In creating measures, the Receptiviti API uses samples that exceed 350 words for baselining its measures (https://docs.receptiviti.com/). Based on this, we impose minimum word limits for our investigation, that is documents that have less than 350 words are too short to be interpreted for our analysis.

Table 1 provides our sample selection criteria and Table 2 provides an overview of the sample composition in terms of industry, region and year.

Table 1: Selection criteria

Step 1: Criteria of selection of companies	Excluded amount	Remaining number of companies
SR for 2016	-	2,649
Region of headquarters - Europe, USA, UK	(1,386)	1,263
Excluding non-EU	(85)	1,178
Excluding healthcare, financials and real-estate sectors and governmental and education services	(293)	885
Excluding companies with missing financial or sustainability data on the database	(306)	579
Manual download of SR	(171)	408
Step 2: Criteria for CEO letter selection		
Expected number of. SR CEO letters in 10 years	-	4,080
Excluding integrated reporting formats	(474)	3,607
Excluding SR with no CEO letters	(390)	3,217
Excluding SR with no CEO letters in English	(16)	3,201
Excluding CEO letters with less than 350 words	(704)	2,496
Excluding SR CEO letter not able to process (e.g., secured pdf files)	(76)	2,420
Excluding SR CEO letter with no sustainability information	(281)	2,139
Available SR CEO letter		2,139
Step 3: Matching FR CEO letters with SR CEO letters		
Expected number of FR CEO letters matching with FR CEO letters	-	2,139
Excluding non-matching FR CEO letters with SR CEO letters	(264)	1,875
Excluding FR with no CEO letters	(526)	1,349
Excluding FR CEO letters with less than 350 words	(15)	1,334
Excluding SR CEO letter with no financial information	(93)	1,241
Excluding FR CEO letters not able to process (e.g., secured pdf files)	(5)	1,236

Remaining number of companies	Matching CEO letters
291	1,236

Table 2: Sample selection

Panel A: Industry composition	Number of observations	% of observations
Communication services	77	6.23
Consumer discretionary	146	11.81
Consumer staples	253	20.47
Energy	145	11.73
Industrials	285	23.06
Information technology	20	1.62
Materials	213	17.23
Utilities	97	7.85
Total	1,236	100.00

Panel B: Region composition	Number of observations	% of observations
EU	391	31.63
UK	262	21.20
USA	583	47.17
Total	1,236	100.00

Panel C: Year composition	Number of observations	% of observations
2010	67	5.42
2011	88	7.12
2012	105	8.50
2013	127	10.28
2014	127	10.28
2015	126	10.19
2016	140	11.36
2017	136	10.96
2018	157	12.70
2019	163	13.19
Total	1,236	100

As for the control variables in our analyses (cf. infra), we collect firms' financial and SP indicators and governance measures from Thomson Reuters' Asset4ESG database and Bureau van Dijk's Orbis database. Thomson Reuters' Asset4ESG database provides standardized information on firms' SP (Dorfleitner et al., 2020; Hawn & Ioannou, 2016; Ioannou & Serafeim, 2012; Shahbaz et al., 2020; Utz, 2017)[6].

3.2. Methods

To test our hypotheses, we employ the following multivariate regression models:

> Linguistic style = *f* [reporting genre, region dummies, financial performance, sustainability performance, media exposure, governance controls, industry dummies, year dummies]

3.2.1. Dependent variables

We focus on several linguistic devices to examine the rhetorical appeals of top management through a more generic approach. Our selection of linguistic devices mainly relies on prior literature that compare and contrast linguistic devices between FR and SR (i.e., Fuoli, 2018; Skulstad, 2005). Skulstad's (2005) comparison in language in FR and SR relies on rhetorical devices that reveal the strategic nature of corporate narrative reporting and communication, such as how managers attempt to persuade and for what purpose (i.e., metadiscourse)[7] (Hossain et al., 2019; Tausczik & Pennebaker, 2010). Specifically, the use of verbs, prepositions, conjunctions and adverbs in FR and SR introductions are analyzed in the study of Skulstad (2005). Such textual devices are considered "function words", which help provide shortcuts, structure and context to communication through the use of typically short and common words, including articles (e.g., "*a*", "*an*", "*the*"), prepositions (e.g., "*to*", "*for*",

---

[6] For example, the environmental score of a firm consists of various factors, such as energy used, water and waste recycled, CO2 emissions; and the social score consists of factors such as employee turnover, injury rate and women employees (Hawn & Ioannou, 2016).

[7] Metadiscourse is an important way of understanding language in written texts which can help enlighten the differences between different genres (Hyland, 2005, p. 31). With metadiscourse, writers construct a discourse according to readers' demands and needs (Hyland, 1998a, 1998b; 2005, p. 71). It allows managers to convey their attitudes and stance towards text content and create a writer-reader interaction (Hyland, 2005, p. 88), and influence audience perception by moderating the persuasiveness of the message and by adding a sense-giving layer (Brennan & Merkl-Davies, 2014; Im et al., 2021). Managers, thus, construct a "reality" to convince people of their chosen or preferred point of view regarding organizational actions, policies or performances in response to their target readers and their shared conventions and expectations (Burke, 1969; Higgins & Walker, 2012).

"*of*"), pronouns (e.g., "*I*", "*them*", "*we*"), auxiliary verbs (e.g., "*was*", "*have*", "*am*"), conjunctions (e.g., "*and*", "*but*", "*since*"), negations (e.g., "*no*", "*not*", "*never*"), and nonreferential adverbs (e.g., "*so*", "*really*", "*very*") (Pennebaker et al., 2014; Boyd et al., 2022). They differ from content words and reflect how people speak (language style), whereas content words carry meanings on their own and reflect what people say (Tausczik & Pennebaker, 2010). The use of content words tends to be context-specific, while the use of function words is to a large extent independent from contexts (Chung & Pennebaker, 2007). Function words are style-related but hold little semantic content outside the context of a sentence. Thinking and attentional patterns are to a large extent reflected in the use of function words (Tausczik & Pennebaker, 2010). Different patterns of function words have been found to be related to overarching thinking styles, honesty and self-focus, status and power, and emotional tone (e.g., Boyd et al. 2022). With such rhetorical devices managers enhance their storytelling in their letters (Martins et al., 2019). Similarly, Fuoli (2018) compares stance expressions as linguistic markers in FR and SR to explore how managers build trust with readers. Stance expressions reflect personal feelings and attitudes (i.e., attitudinal stance), certainty and reliability (i.e., epistemic stance), and permission, ability and obligation (i.e., modality). Based on these categories, he retrieves the frequencies of grammatical markers (e.g., *desire/intention/decision verbs* or *ability and willingness adjectives* for attitudinal stance, *certainty and likelihood adverbs* for epistemic stance, *necessity and permission modals* for modality). According to Fuoli (2018), the frequent use of adverbs such as "*actually*" or "*obviously*", modals such as "*may*" or "*could*", and verbs such as "*intend*", "*believe*" or "*expect*" reflect rationality, cautiousness and competence as more observed in FR; and frequent use of adverbs and adjectives such as "*amazingly*" or "*afraid*", modals such as "*can*" or "*must*", and verbs such as "*aim*", "*show*" or "*understand*" reflect a more committed, affective and honest language, as in SR. The adverbs, modals and verbs that Fuoli (2018) studies are considered cognitive markers.

They relate to the emotional and cognitive dimensions of metadiscourse and affect the persuasiveness and credibility of messages (Aerts & Yan, 2017). In light of these studies, we use the linguistic devices: (1) quantification; (2) time orientation; (3) causality; (4) cognitive marking mechanisms; (5) personalization; and (6) affective processes markers. The selected linguistic devices are shown to direct readers' understanding of the information presented and to examine the preparer's stance to the message that goes beyond thematic and propositional content. To capture the linguistic devices in FR and SR, we follow a dictionary approach using text analysis software, LIWC-22 (Boyd et al., 2022). LIWC-22 measures verbal variability and context mainly based on word frequencies that relate to psychosocial constructs and theories (Boyd et al., 2022).

First, we use quantification, time orientation and causal language markers. Prior research shows that quantitative communication indicates a stronger commitment to specific managerial objectives, providing more precise information with data (i.e., hard information), as well as a better quality of disclosures (Dyer et al., 2017; Silva, 2021; Stacchezzini et al., 2016). To measure quantitative communication, we use numbers (e.g., "*one", "two", "first", "once"*) and quantity (e.g., "*all", "one", "more", "some"*) dictionaries of LIWC-22. Next, we use time orientation markers of LIWC-22 (e.g., "*when", "now", "day", "year"*), as messages in narrative disclosures can be constructed in both a backward- or a forward-looking manner, which relates to providing measurable, accountable and concrete data and evidence (Silva, 2021). We use conjunctions (e.g., "*and*", "*but*", "*so*", "*as*") and causation (e.g., "*how*", "*because*", "*make*", "*why*") to measure attributions in CEO letters, which are used to link corporate events or performance outcomes to a reason or a cause (Aerts, 2005; Zhao et al., 2016). In addition, differentiation words (e.g., "*but", "if", "else"*) are used to measure comparisons in texts, that relate to the factual stance (Hyland, 2005). We expect these markers to reflect the descriptive, factual and rational language style owing to their quantitative, time-oriented, explanatory and performance comparison indications.

We use cognitive marking mechanisms: insight (e.g., *"think", "know", "feel"*) and certitude (e.g., "*really", "actually", "of course"*) words that help convey an impression of certainty (Boyd et al., 2022) that include boosters; and discrepancy (e.g., "*should*", "*would"*) and tentative words (e.g., "*maybe*", "*perhaps"*) which consist of hedges that are used to realize the cautious, honest and sincere appeal (Hyland, 2005). Finally, we measure affective processes words that consist of, positive/negative tone and emotion words (e.g., *"good", "happy", "bad", "wrong"*) that reflect both sentiment and emotion (Boyd et al., 2022), and "*I"* and other-pronouns (e.g., *"I", "me" and "we", "you"*) that reflect self-serving IM and authoritative stance (Garzone, 2005; Hyland, 2005) and collectivism and solidarity (Harjoto et al., 2021), respectively. Anticipated markers are expected to manifest a more informal and conversational storytelling style due to the cognitive and affective indication mechanisms.

We take the natural logarithm of the identified linguistic devices and employ factor analysis with varimax orthogonal rotation that can proxy managers' stances. The result of the factor analysis shows three uncorrelated factors with eigenvalues greater than 1.0 which cumulatively explain 55% of the overall variance, of which the latter two factors relate to different motives of the storytelling stance of top management. Table 3 shows the factor loadings.

Table 3: Rotated factor loadings

Variable	Examples	F-1	F-2	F-3
Numbers	billion, first, second, quarter	0.29	-0.22	-0.73
Quantities	amount, average, less, majority	0.89	-0.01	-0.00
Time	ago, continue, daily, regularly	0.76	0.03	0.24
Conjunctions	because, so, when, whilst	0.51	0.01	0.71
Insight	believe, define, explore, know	0.08	0.32	0.60
Causation	depend, effect, outcome, react	0.11	0.07	0.69
Discrepancy	can, could, want, would	-0.05	0.57	0.35

Tentative	assume, hope, might, perhaps	0.41	0.54	-0.14
Certitude	clear, directly, every, specific	0.06	0.69	0.01
Differentiation	apart, exclude, split, unlike	0.61	0.39	0.17
"*I*" pronouns	I, me, mine, my	-0.11	0.46	0.06
Other pronouns	her, him, their, we	0.06	-0.04	0.69
Affect	afraid, benefit, disaster, luck	0.30	-0.10	0.73

For Factor 1, we observe high loadings for numbers, quantities, time, conjunctions and differentiation, a pattern that reflects a stance based on more quantitative, factual and time-based logical thinking. This approach tends to be highly descriptive and somewhat impersonal, in which commitment to facts and concrete outcomes is high. In addition, we observe a relatively high load of tentative words which may be used to decrease the risk of commenting inaccurately on organizational achievements (Aerts & Yan, 2017). We also highlight the presence of affective language in our Factor 1. For Factor 2, we observe high loadings for a range of cognitive processes indicators, such as discrepancy, tentative, differentiation and certitude. These indicators include both boosters and hedges, which Hyland (2005) refers to as "*double-edged*." Hyland (2005, pp. 81) argues that "*boosters allow writers to project a credible image of authority, decisiveness and conviction, while hedges demonstrate personal honesty and integrity through a willingness to address hard realities, albeit behind a shield of mitigation".* In addition, Factor 2 strongly loads on *"I"* pronouns that indicate the CEO's personal commitment. Hedges, boosters and self-mentions are commonly used to realize a competent and honest stance to project an aura of credibility gained by openness and tolerance for ambiguity (Aerts & Yan, 2017; Hyland, 1998). In that sense, Factor 2 tends to project an image of a modest, trustworthy and cautious steward or leader of the company. Unlike Factor 1, Factor 2 is more personal and informal rather than descriptive with much less commitment to concrete facts and circumstances. Remarkably, Factor 2 loads indifferent on affective language, pointing to a tendency to use neutral language in commenting on organizational events. Finally, in Factor 3, we observe dominant loadings on the use of 'other pronouns' and affective language (including both positive and negative emotion words). The 'other pronouns' category is instrumental in building a storytelling atmosphere based on indicators of stakeholder commitment and solidarity (Araujo & Kollat, 2018; Harjoto et al., 2021). This factor also scores high on conjunctions and causation words, which would promote the build-up of stories through longer, causal and interdependent statements. In addition, the frequent use of insight words helps to reconstrue or reframe events in a more emotional way (Pennebaker et al., 1997; Tausczik & Pennebaker, 2010)[8], which is in line with the affective component of this factor. In our data, Factor 3 also shows to be highly correlated with positive (net) tone, indicating that the 'affect' component tends to stand for positive affective language. Considering these components, we use Factor 3 as a proxy for an assertive relationship-oriented stance. Whereas Factor 1 uses a detached, factual and rational approach to add meaning and context, Factor 2 and Factor 3 are more oriented towards building a story, although through a different stance. Factor 2 builds on first-person-centered commenting with high cognitive commitment, although its double-edged nature and neutral wording make it less prone to opportunistic assertive IM. Factor 3 tends to invest in relationship framing with positive tone management in an informal, more conversational style. Based on these, we label the factors as (1) rational-factual; (2) personalized sense-giving; and (3) assertive relationship-oriented.

[8] Pennebaker et al. (1997) studied the words people use in disclosing a trauma to predict improvements in mental and physical health. The study shows that when people use more of causal and insight words in their writing about past events, they tend to feel better afterwards. This is because using these types of words shows that we are actively thinking about and processing the event, which can help us make sense of it. Insight words show that we are actively thinking about our own thoughts and feelings.

3.2.2. Independent variable and controls

Our variable of interest is reporting genre (*SR*) (H1a and H1b), which is a dummy variable that is equal to 1 for SR and 0 for FR. We test the moderation effect of regions to further test our regional-level institutional setting hypotheses (H2a and H2b) using a region dummy (*US_dum*), that is equal to 1 if company is from the US; 0 otherwise. Controls include key financial and SP variables, as rhetorical IM is investigated through the relationship between the financial or SP of firms and linguistic indicators of interest (Boiral et al., 2020; Dunne et al., 2021; Smeuninx et al., 2020; Stacchezzini et al., 2016). In our models, we gradually add control variables to observe any change in results. The financial performance controls include return on assets (*ROA*) and return on equity (*ROE*). We include a loss dummy variable (*LOSS*), which equals 1 if the company has negative net income, and 0 otherwise. As companies' financial risk can influence management's disclosure practices (Cormier & Magnan, 2003; Orlitzki & Benjamin, 2001), we include financial leverage using the debt-to-equity ratio (*LEV*) and liquidity risk using the current ratio (*LIQ*). Previous studies show that both company size and firm value are significantly associated with the quantity and quality of information disclosed in FR and SR (e.g., Aerts & Yan, 2017; Aouadi & Marsat, 2018; Clarkson et al., 2008; Fiandrino & Tonelli, 2021; Mittelbach-Hörmanseder et al., 2020). We measure company size (*SIZE*) as the natural logarithm of total assets, and firm value using Tobin's Q (*TOBQ)*, calculated as firms' market capitalization scaled by total assets.

Following previous studies (e.g., Braam & Peeters, 2018; Ioannou & Serafeim, 2012; Michelon et al., 2015), we control for SP using the mean of the environmental and social pillar scores (*ENVSOC)* provided by Asset4ESG database. The environmental score includes measures for factors like resource and emission reduction and product innovation, and the social score measures items related to employee health and safety, human rights and diversity (Ribando & Bonne, 2010). Previous studies have shown that higher levels of media exposure increase public scrutiny regarding the SP of firms, which significantly affects SR disclosure (Aerts & Cormier, 2009). Corporate controversies are events that can be publicly observed through media, which impact firms' reputations and legitimacy negatively (Del Giudice & Rigamonti, 2020; Utz, 2017). Thomson Reuter's Asset4ESG database provides an ESG controversies score, which measures environmental and social scandals reflected in the media using different major English-speaking news and non-governmental organizations (NGOs) (Aouadi & Marsat, 2018; DasGupta, 2021; Shakil, 2021). We implement the controversies score as a dummy variable in our models where 1 stands for firms with controversies; 0 otherwise (*ESG_contro*). As argued by Aouadi and Marsat (2018) controversies can enhance firm visibility, and thus, affect firm value. In addition, corporate governance characteristics such as board communication and overlapping directorship, and expertise in sustainability at the board level are shown to affect FR and SR quality (Birkey et al., 2016; Moses et al., 2020; Shakil, 2021). Therefore, we include board size (*BOARD*_size), CEO-duality (dummy) (*CEO_duality*) and CSR committee availability (dummy) (CSR_com) in our model. Moreover, we include industry and year indicators. Prior research has established that litigation risk and environmentally sensitive activities may affect the use of language in FR and SR, which can be proxied by industry indicators (Aerts & Cormier, 2009; Al Hawaj & Buallay, 2021; Cho & Roberts, 2010; García-Meca & Martínez-Ferrero, 2021). We use the Global Industry Classification Standard (GICS) for industry classification.

We provide variable definitions in Table 4.

Table 4: Variable definition

Variable	Definition	Source
Rational-factual	Predicted rational-factual language scores based on factor analysis	Authors' calcul.
Personalized sense-giving	Predicted personalized sense-giving language scores based on factor analysis	Authors' calcul.
Assertive relationship-oriented	Predicted assertive relationship-oriented language scores based on factor analysis	Authors' calcul.
SR	Dummy variable 1 if SR, 0 if FR	Websites

ROA	Net income / average total assets	Thomson Reuters
ROE	Net income / average total equity	Thomson Reuters
LOSS	Loss company dummy: 1 if loss (net income < 0); 0 otherwise	Authors' calcul.
LEV	Leverage indicator (debt-to-equity ratio): Total debt / shareholders equity	Thomson Reuters
LIQ	Liquidity indicator (current ratio): Total current assets / total current liabilities	Thomson Reuters
SIZE	Natural logarithm of total assets	Authors' calcul.
TOBQ	Tobin's Q: Market capitalization / total assets	Bureau van Dijk
ENVSOC	The average of the environmental and social scores in the Asset4ESG database	Thomson Reuters
ESG_contro	ESG controversies dummy: 1 if the company has a controversies score less than 100; 0 otherwise	Authors' calcul.
Board_size	Number of board members	Thomson Reuters
CEO_duality	Dummy variable: 1 if CEO and chairman are the same; 0 otherwise	Thomson Reuters
CSR_com	CSR committee dummy: 1 if yes; 0 otherwise	Thomson Reuters

4. Results

4.1. Descriptive statistics

To limit the influence of outliers, we winsorize our continuous variables at the top and bottom tails at a 1% level. To address any potential multicollinearity problem, we examine the variance inflation factors (VIF) of the data. As they are all below the recommended thresholds (<10) (Hair et al., 2010), our models indicate no multicollinearity issues between variables. Table 5 summarizes the descriptive statistics of the variables used in the regression models. Table 6 presents the Pearson correlation coefficients.

As descriptives show, almost half of our observations are from the US (47%), which provides a robust dataset to test the effect of regional-level institutional settings on the use of rhetorical appeals. The mean values for our financial controls are 0.59 for ROA and 0.23 for ROE. The mean of LOSS shows that only 10% of companies in our dataset relate to loss companies. Descriptives show that generally, our dataset consists of companies that have slightly better ENVSOC scores (μ = 67.31). ESG_contro mean is noted as 0.44, which shows that more than half of the companies in our dataset did not have any ESG controversy in the respected firm-year observation. Regarding our corporate governance variables, we observe that companies in our dataset have 11.39 board members, on average; and mainly CEO-chairman separation does not take place (μ = 0.41). This suggests that CEO duality in our sample is slightly higher. In our sample, 91% of the companies included have a separate CSR committee.

Correlations coefficients indicate that SR is negatively correlated with rational-factual language (r = -0.370) and positively correlated with assertive relationship-oriented (r = 0.440), providing initial support for H1 and H2. Regarding the other storytelling appeal that is identified through factor analysis, personalized language, the sign of the correlation coefficient is in line with our expectation, yet the correlation is low (r = 0.006). The correlation analysis shows that the US_dum is positively correlated with the rational-factual language (r =0.082) and assertive-relationship oriented language (r = 0.099), and negatively correlated with the personalized sense-giving language (r = -0.221).

Next, we carry out additional analyses utilizing multivariate regressions.

Table 5: Descriptive statistics

	Obs.	Mean	SD	Min	Max	VIF
Rational-factual	2,472	0	1	-17.3	2.4	
Personalized sense-giving	2,472	0	1	-4.72	2.88	
Assertive relationship orient.	2,472	0	1	-14.85	2.78	
SR	2,472	0.5	0.5	0	1	1.89
US_dum	2,472	0.47	0.50	0	1	2.73
ROA	2,472	0.59	0.64	-1.43	2.74	2.78
ROE	2,472	0.23	0.38	-1.35	2.19	1.35
LOSS	2,472	0.1	0.29	0	1	1.26
LEV	2,472	1.25	1.67	-0.03	12.45	1.26

LIQ	2,472	1.36	0.63	0.45	3.75	1.54
SIZE	2,472	23.46	1.51	20.18	26.72	2.74
TOBQ	2,472	1.17	1.02	0.13	5.01	2.84
ENVSOC	2,472	67.31	16.55	24.04	94.03	1.68
ESG_contro	2,472	0.44	0.5	0	1	1.50
Board_size	2,472	11.39	3.21	1	23	1.40
CEO_duality	2,472	0.41	0.49	0	1	1.59
CSR_com	2,472	0.91	0.29	0	1	1.15

Table 6: Pairwise correlations

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
(1) Rational-factual	1.000																
(2) Personalized sense-giving	0.000	1.000															
(3) Assertive relationship.	0.000	0.000	1.000														
(4) SR	-0.370*	0.006	0.440*	1.000													
(5) US_dum	0.082*	-0.221*	0.099*	0.000	1.000												
(6) ROA	-0.037	-0.044	0.010	-0.001	-0.009	1.000											
(7) ROE	-0.046	-0.006	-0.019	0.000	0.047	0.334*	1.000										
(8) LOSS	0.011	0.027	-0.022	0.001	-0.057*	-0.344*	-0.140*	1.000									
(9) LEV	-0.043	-0.051	-0.040	-0.001	0.069*	-0.026	0.337*	0.078*	1.000								
(10) LIQ	-0.036	-0.054*	0.068*	0.000	0.123*	0.146*	-0.033	0.060*	-0.129*	1.000							
(11) SIZE	0.082*	0.044	-0.018	0.000	0.176*	-0.219*	-0.108*	-0.031	0.023	-0.304*	1.000						
(12) TOBQ	-0.061	-0.080*	0.046	0.000	0.047	0.733*	0.301*	-0.188*	-0.019	0.169*	-0.350*	1.000					
(13) ENVSOC	0.019	0.088*	0.018	0.000	-0.144*	0.047	-0.018	-0.073*	-0.070*	-0.195*	0.449*	0.002	1.000				
(14) ESG_contro	0.063*	0.070*	-0.002	0.001	0.064*	-0.057*	-0.059	-0.010	-0.043	-0.149*	0.523*	-0.105*	0.286*	1.000			
(15) Board_size	0.044	0.052	-0.020	0.000	0.027	-0.119*	-0.082*	-0.037	-0.001	-0.137*	0.474*	-0.180*	0.302*	0.284*	1.000		
(16) CEO_duality	0.077*	-0.151*	0.025	0.000	0.547*	0.004	-0.007	-0.074*	-0.010	-0.024	0.191*	0.049	-0.020	0.115*	0.186*	1.000	
(17) CSR_com	-0.004	0.031	-0.039	0.000	-0.032	-0.007	0.027	-0.031	0.010	-0.107*	0.188*	-0.032	0.299*	0.104*	0.097*	-0.049	1.000

* Correlations significant at p<0.01

4.2. Multivariate regression analyses

We use a pooled OLS regression model to test our hypotheses. The results of our multivariate regression models are presented in Table 7. The first three models show the direct relationship between our dependent and independent variables, without the interaction variable. Model 1 shows that SR exhibits significantly less use of the rational-factual language than FR (b = -0.741), which is in line with our expectations. In Model 2, we test disclosure genre (SR) as a determinant of the personalized sense-giving language, yet we cannot observe any significant effect ($p > 0.1$). Model 3 demonstrates that top management in SR CEO letters uses the assertive relationship-oriented language significantly more (b = 0.880). These results provide initial support for our hypotheses (H1a and H1b), yet we cannot observe any significant effect of the disclosure genre on the use of personalized sense-giving language ($p > 0.1$), counterintuitively. In the next three models (Model 4, Model 5, and Model 6), we examine the interaction effect of disclosure genre (SR) and companies' region (US_dum) to test the moderation effect (H2a and H2b). The results support our expectations, which predict that the US institutional setting significantly moderates the use of all rhetorical appeals identified, and adding the interaction term in the Models (4-6) significantly improves model fit, i.e., the interaction terms are statistically significant and R^2 increases. Our results show that the directions of moderation for rational-factual, personalized sense-giving and assertive relationship-oriented are positive, negative and positive, respectively (b = 0.169; b = -0.342; b = 0.146). While we did not formulate any hypothesis regarding the moderation effect of regions on the rational-factual style, our Model 4 demonstrates a positive interaction effect, implying that accountability and transparency of corporate disclosures may still be ahead of the European counterparts. This may relate to the fact that for US companies, stock markets are their most important source of capital (Matten & Moon, 2008); and rational-factual appeal may be used as a response to the scrutiny for information and transparency concerns, also implying a more pragmatic legitimacy concerns. Our Model 5 shows that the use of personalized sense-giving appeal significantly depends on whether or not the company is from the US. The US institutional setting negatively affects the use of personalized sense-giving appeal in SR CEO letters. This result is counterintuitive, yet the rational communication in the US (as Model 4 shows the positive interaction effect on the use of rational-factual language) could limit the use of personalized appeals. We also observe that adding this interaction term in our regression model also makes the main effect significant. As this result alone does not provide much about whether consensual relationship-building is less in US SR CEO letters, further investigation regarding our hypothesis (H2a) is required (see Section 5.3. Additional Analyses). Our Model 6, however, shows the positive relationship between the US SR CEO letters and assertive relationship-oriented language (b = 0.146), which is in line with the explicit institutional setting culture and supports our hypothesis (H2b). We also highlight that this factor includes both "assertive" and "relationship-orientation"; hence, we further investigate it in the next section, as well. When we add controls in the next models (Model 7, Model 8 and Model 9), our results remain robust for all appeals. Moreover, our results with controls show that ESG controversies positively affect the use of rational-factual and personalized sense-giving appeals, which indicates that in case of an ESG controversy, managers tend to use either rational-factual or personalized sense-giving appeals as a reaction. Controversies, that are publicly observed through media, therefore, may restrict managers to adopt assertive relationship-oriented appeal. Figure 1 depicts the interaction effects.

As additional robustness checks, we introduced a GRI dummy variable into the models to signify whether the company adheres to GRI standards and guidelines in its disclosure practices (Chelli et al., 2018). The main results remain consistent, and GRI dummy variable was not significant across all models. However, due to lack of data that leads to a loss of 510 observations, we opted not to include this variable in our analyses. Furthermore, recognizing the unique regulatory environment of extractive industries, which often face additional regulations, such as anti-bribery or anti-corruption measures (Öge, 2016), we conducted supplementary analyses by excluding observations from these industries. The (untabulated) results remained consistent with our main findings. Finally, adjustments for heteroscedasticity and autocorrelation are also implemented through clustering standard errors at the firm-year level. The results are consistent with the presented in Table 7.

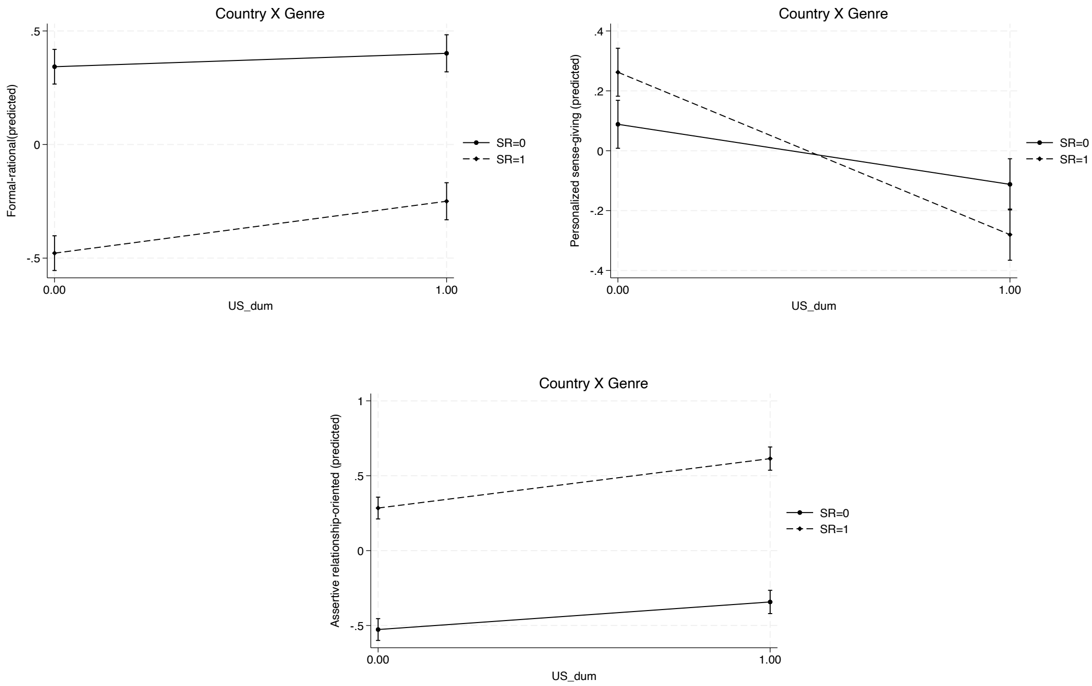

Figure 1: The effect of Genre (SR) and Region (US_dum) on the predicted linguistic appeals. Both lines serve the purpose of highlighting the relative differences based on the binary category (US_dum).

Table 7: Regression results

VARIABLES	(1) Ration. fact.	(2) Personal. sense-giv.	(3) Assert. relat.	(4) Ration. fact.	(5) Personal. sense-giv.	(6) Assert. relat.	(7) Ration. fact.	(8) Personal. sense-giv.	(9) Assert. relat.
SR	-0.741***	0.013	0.880***	-0.821***	0.174***	0.812***	-0.821***	0.174***	0.812***
	(0.037)	(0.039)	(0.035)	(0.051)	(0.054)	(0.049)	(0.051)	(0.053)	(0.049)
US_dum	0.150***	-0.447***	0.229***	0.065	-0.276***	0.156***	0.059	-0.201***	0.184***
	(0.041)	(0.043)	(0.039)	(0.056)	(0.058)	(0.053)	(0.061)	(0.064)	(0.059)
SR X US_dum				0.169**	-0.342***	0.146**	0.169***	-0.341***	0.146**
				(0.074)	(0.078)	(0.071)	(0.074)	(0.077)	(0.071)
ROA	-0.001	-0.008**	-0.007**	-0.001	-0.008**	-0.007**	0.009*	-0.004	-0.006
	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)	(0.005)	(0.005)	(0.005)
LOSS	0.028	0.036	-0.098	0.028	0.035	-0.098	0.060	0.062	-0.087
	(0.070)	(0.074)	(0.067)	(0.070)	(0.073)	(0.067)	(0.071)	(0.074)	(0.067)
LIQ	-0.036	0.029	0.069**	-0.036	0.029	0.069**	-0.048	0.019	0.053
	(0.036)	(0.038)	(0.035)	(0.036)	(0.038)	(0.035)	(0.037)	(0.038)	(0.035)
SIZE	0.026	0.033*	-0.019	0.026	0.033*	-0.019	0.007	-0.008	-0.012
	(0.017)	(0.018)	(0.016)	(0.017)	(0.018)	(0.016)	(0.020)	(0.021)	(0.019)
ENVSOC	0.001	0.002	0.003*	0.001	0.002	0.003*	0.001	0.002	0.003**
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
ESG_contro							0.091**	0.093**	0.000
							(0.046)	(0.048)	(0.043)
ROE							-0.051	0.147**	-0.067
							(0.056)	(0.059)	(0.053)
LEV							-0.027**	-0.032**	-0.026**

							(0.012)	(0.013)	(0.012)
TOBQ							-0.085***	-0.049	-0.005
							(0.031)	(0.032)	(0.029)
Board_size							0.003	0.008	-0.004
							(0.007)	(0.007)	(0.006)
CEO_duality							0.069	-0.118**	-0.034
							(0.047)	(0.049)	(0.045)
CSR_com							-0.044	0.020	-0.117*
							(0.069)	(0.072)	(0.065)
Industry dummies	Included	Included	Included	Included	Included	Included	Included	Included	Included
Year dummies	Included	Included	Included	Included	Included	Included	Included	Included	Included
Constant	-0.186	-0.500	-0.310	-0.146	-0.581	-0.276	0.604	-0.042	-0.252
	(0.399)	(0.419)	(0.379)	(0.399)	(0.418)	(0.379)	(0.456)	(0.477)	(0.433)
Observations	2,472	2,472	2,472	2,472	2,472	2,472	2,472	2,472	2,472
R-squared	0.156	0.068	0.238	0.157	0.076	0.239	0.165	0.083	0.243

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Furthermore, as our regional-level institutional setting hypotheses (H2a and H2b) consider only SR in the context of storytelling, we run an additional analysis dropping FR observations. The results are shown in Table 8, which are in line with our main results (cf. supra Table 7): showing a positive effect of the US institutional effect on the use of rational-factual language, negative effect on personalized sense-giving language and positive effect on assertive relationship-oriented language.

Table 8: Regression results

VARIABLES	(1) Ration. fact.	(2) Personal. sense-giv.	(3) Assert. relat.	(4) Ration. fact.	(5) Personal. sense-giv.	(6) Assert. relat.
US_dum	0.213***	-0.611***	0.320***	0.187***	-0.489***	0.328***
	(0.0601)	(0.0662)	(0.0511)	(0.0711)	(0.0782)	(0.0604)
ROA	-0.00325	0.00212	0.00359	0.00602	0.00461	-0.00514
	(0.00506)	(0.00558)	(0.00430)	(0.00700)	(0.00770)	(0.00595)
LOSS	0.0529	0.0884	0.00368	0.0825	0.0972	-0.0301
	(0.102)	(0.112)	(0.0867)	(0.103)	(0.114)	(0.0877)
LIQ	-0.00870	0.0414	-0.00946	-0.0115	0.0284	-0.00333
	(0.0530)	(0.0584)	(0.0451)	(0.0537)	(0.0591)	(0.0456)
SIZE	0.0579**	0.0337	-0.0138	0.0202	0.0221	0.0157
	(0.0244)	(0.0269)	(0.0207)	(0.0295)	(0.0325)	(0.0251)
ENVSOC	0.0045**	0.00431*	0.00193	0.00511**	0.00429*	0.00239
	(0.00201)	(0.00222)	(0.00171)	(0.00211)	(0.00233)	(0.00180)
ESG_contro				0.0880	0.110	-0.00938
				(0.0666)	(0.0733)	(0.0566)
ROE				0.0107	0.116	-0.00394
				(0.0819)	(0.0901)	(0.0696)
LEV				-0.0200	-0.0241	0.0171
				(0.0179)	(0.0197)	(0.0153)
TOBQ				-0.0856*	-0.0397	0.0735*
				(0.0446)	(0.0490)	(0.0379)
Board_size				0.00910	-0.00780	-0.0172**
				(0.00995)	(0.0109)	(0.00845)
CEO_duality				0.0993	-0.217***	-0.0562
				(0.0691)	(0.0760)	(0.0587)
CSR_com				-0.125	0.0378	-0.0587
				(0.100)	(0.110)	(0.0852)
Industry dummies	Included	Included	Included	Included	Included	Included

Year dummies	Included	Included	Included	Included	Included	Included
Constant	-2.049***	-0.382	0.620	-1.213*	-0.0431	0.127
	(0.580)	(0.640)	(0.640)	(0.664)	(0.730)	(0.564)
Observations	1,236	1,236	1,236	1,236	1,236	1,236
R-squared	0.057	0.108	0.076	0.066	0.119	0.084

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

4.3. Additional analyses

In an effort to further investigate/explore the meaning and ramifications of the rhetorical style factors that we elaborated, we use preset summary variables of LIWC (analytical thinking, clout and emotional tone). These summary variables are considered generic, i.e., not genre-specific, hence useful to measure overall tone in a mixed sample consisting of both FR and SR. They allow us to disentangle and unravel the composite Factor 3 (assertive relationship-oriented) which thrives on two narrative mechanisms, relationship-framing and positive tone management which may interact differently with institutional norms and context. It also allows us to trim down on the affective side of our Factor 1 (rational-factual).

The summary variables of LIWC are scored by proprietary algorithms constructed from large text samples used in previous research conducted by the developers of the software (Boyd et al., 2022). They are converted to percentiles based on standardized scores derived from large comparison samples of modern texts (Boyd, 2017). Although the proprietary algorithms are to a certain extent non-transparent, the language categories that are algorithmically nested under the summary variables have been well-documented. The summary variables are composite metrics and mainly rely on the occurrence of function words in the narratives. A key LIWC summary metric that we will use as an alternative to our Factor 1 (rational-factual) is referred to as ‘analytic’ or ‘analytical thinking’. The construct of analytical thinking refers to a deliberate mode of reasoning whereby complex concepts are deconstructed into more manageable parts and their interrelations. Analytical thinking transpires in verbal behavior through the use of articles, which signal concepts, and prepositions, which convey relationships between concepts (Davies, 2011; Jordan et al., 2019). The construct allows us to differentiate information-structural language versus story-driving language. Texts high in analytical thinking tend to be formal, logical, and hierarchically structured in terms of how information is presented, while those low in analytic thinking are more prone to include personal stories, more action of agents and events, and a sense of immediacy (Pennebaker et al., 2015; Tay, 2021). The latter's focus on people and actions, indicating a more informal, experiential and intuitive way of communicating ideas and actions. Such an informal type of thinking can be measured in language through the use of pronouns, adverbs, negations, auxiliary verbs, and conjunctions (Pennebaker et al., 2015; Pennebaker et al., 2014)[9]. The distinction is, however, not typological as all texts contain elements of both (Fuoli, 2018). CEO letters have both an analytic and a narrative component, but they are likely to differ on a continuum along which information-structural language versus story-driving language is used. The summary measure is conceptually near to our Factor 1 (rational-factual), but without the link to affective language as a measurement component. In that sense, the variable 'analytic' captures less emotion as a rhetorical persuasion mechanism.

A second LIWC summary measure, 'Clout', captures a language style gauged by the presence of authoritative, confident and outward-facing language that reflects leadership, certainty and being in control (Kacewicz et al., 2014; Pennebaker et al., 2015; Tay, 2021). The clout variable is derived from prior research focusing on personal and social interaction and is highly grounded in pronoun use which is significantly correlated with social hierarchy (Boyd et al., 2022; Kacewicz et al., 2014). Higher clout score is marked by using more we-words, you-words and social words and fewer I-words, negations (e.g., "*no*", "*not*"), swear and differentiation words. Referring to collective or others' work and interests or how one is connected with other reference groups, will have a higher clout score as the text would use more "*we*" and "*you*" pronouns, and social words compared with texts with high self-focus and low social interaction using "*I*" first-person singular pronouns (Adaji & Olakanmi, 2019;

[9] Analytic thinking is calculated as follows before being normalized, [articles + prepositions - pronouns - auxiliary verbs - adverb - conjunctions - negations].

Duncan et al., 2019; Oliver et al., 2020). Using more outward-facing words and inclusive language, writers demonstrate that they focus more on the people they are interacting with versus themselves (Adaji & Olakanmi, 2019). This type of interaction would be well-suited in a communication format that aims to foster relationship-building and engagement as part of knowledge construction in the diffuse sustainability field. Focusing on relationship building is impounded as one of the two most important language categories in our Factor 3 (assertive relationship-oriented). Using social and inclusive language conveys authority and confidence in the company's future, leading the audience to form a preferred impression (Sparks & Areni, 2007; Toma & D'Angelo, 2015).

The second most important component of our Factor 3 (assertive relationship-oriented) relates to the use of affective words, which is typically captured by the third LIWC summary variable, 'emotional tone', that we will use in this study. The variable emotional tone reflects the degree of positive emotional tone, as captured by the LIWC language categories *affect*, *positive* and *negative emotion* (Cohn et al., 2014; Pennebaker et al., 2015; Tay, 2021). The variable entails emotions that have arousing effects which tend to drive attention and attitude change (Petty & Cacioppo, 1986). Using both clout and emotional tone as separate style variables helps us to disentangle the ramifications of the composite Factor 3 (assertive relationship-oriented) that we derived from our factor analysis.

Table 9 presents the correlations between the summary variables, factors that we identified and linguistic indicators.

Table 9: Pairwise correlations

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
(1) Analytic	1.000																		
(2) Clout	-0.461*	1.000																	
(3) Tone	-0.285*	0.498*	1.000																
(4) Rational-factual	0.186*	-0.048*	0.034	1.000															
(5) Personalized sense-givi.	-0.239*	-0.125*	-0.180*	0.000	1.000														
(6) Assertive relationship or.	-0.496*	0.605*	0.560*	0.000	0.000	1.000													
(7) Numbers	0.446*	-0.302*	-0.311*	0.287*	-0.217*	-0.729*	1.000												
(8) Quantities	0.172*	-0.003	0.034	0.886*	-0.014	-0.004	0.262*	1.000											
(9) Time	0.070*	0.071*	0.129*	0.760*	0.032	0.243*	0.001	0.592*	1.000										
(10) Conjunctions	-0.279*	0.332*	0.328*	0.507*	0.006	0.712*	-0.403*	0.406*	0.500*	1.000									
(11) Insight	-0.269*	0.212*	0.155*	0.080*	0.317*	0.602*	-0.382*	0.066*	0.201*	0.462*	1.000								
(12) Causation	-0.297*	0.246*	0.204*	0.108*	0.065*	0.686*	-0.406*	0.080*	0.214*	0.504*	0.427*	1.000							
(13) Discrepancy	-0.303*	0.142*	0.104*	-0.054*	0.571*	0.347*	-0.290*	0.019	0.096*	0.187*	0.270*	0.236*	1.000						
(14) Tentative	-0.059*	-0.079*	-0.133*	0.411*	0.541*	-0.137*	0.018	0.253*	0.186*	0.134*	0.106*	0.024	0.150*	1.000					
(15) Certitude	-0.152*	-0.031	-0.049*	0.067*	0.694*	0.009	-0.114*	0.075*	0.127*	0.058*	0.176*	0.069*	0.213*	0.194*	1.000				
(16) Differentiation	-0.086*	-0.054*	-0.004	0.610*	0.390*	0.174*	-0.115*	0.445*	0.363*	0.417*	0.250*	0.225*	0.201*	0.365*	0.199*	1.000			
(17) "I" pronouns	-0.175*	0.071*	0.060*	-0.114*	0.462*	0.064*	-0.080*	-0.027	0.047	-0.019	0.092*	-0.041	0.165*	0.050	0.170*	0.034	1.000		
(18) Other pronouns	-0.523*	0.898*	0.463*	0.062*	-0.036	0.688*	-0.351*	0.095*	0.177*	0.418*	0.285*	0.312*	0.211*	-0.018	0.034	0.104*	0.119*	1.000	
(19) Affect	-0.241*	0.466*	0.781*	0.298*	-0.105*	0.732*	-0.377*	0.258*	0.375*	0.598*	0.306*	0.370*	0.181*	-0.016	0.018	0.217*	0.102*	0.541*	1.000

* Correlations significant at p<0.01

The correlation table shows that identified *analytic* positively correlated (0.186) with our identified rational-factual appeal, and negatively correlated with personalized sense-giving and assertive relationship-oriented appeals (r = -0.239 and r = -0.496, respectively). *Clout* measure is negatively correlated with our identified rational-factual and personalized sense-giving appeals (r = -0.048 and r = -0.125, respectively), and positively correlated with the identified assertive relationship-oriented appeal (r = 0.605). Finally, *tone* is positively correlated with rational-factual and assertive relationship-oriented appeals (r = 0.034 and r = 0.560, respectively) and negatively correlated with personalized sense-giving appeal (r = -0.180). This shows that especially *clout* measure captures both personalized sense-giving and assertive relationship-oriented appeals, but the difference is mainly driven because of the use of other pronouns and affect. While *clout* is mainly captured through other pronouns (r = 0.898), *tone* is mainly captured through the intensity of affective words (r = 0.781).

We further examine our hypotheses using multivariate regression models with LIWC's summary variables (i.e., *analytic*, *clout*, *tone*). The descriptives and correlations of summary variables with independent and control variables are provided in Table 10 and Table 11.

Table 10: Descriptive statistics of LIWC summary variables

	Obs.	Mean	SD	Min	Max
Analytic	2,472	86.28	8.83	21.60	99.00
Clout	2,472	91.24	11.56	27.74	99.00
Tone	2,472	76.54	13.96	13.21	99.00

Table 12 shows the multivariate OLS regression results using the LIWC-22 summary variables. Models 1, 2 and 3 show that the relationship between the main effect of SR and summary variables are significant. Results show that the effect of SR is negative on *analytic* (b = -6.052) and positive on *clout* and *tone* (b = 5.748 and b = 4.084). In Model 4, Model 5 and Model 6, we observe that the interaction effect of the US institutional setting. In Model 4, we observe that the interaction effect is not significant for *analytic*. Yet, we observe that the region effect negatively moderates the use of *clout* (b = -3.333) (Model 5) and positively moderates the use of *tone* (b = 2.801) (Model 6). These results suggest that in Europe top management mainly uses credible and powerful language for storytelling purposes, while in the US tone management is more prominent. The use of credible and powerful language in Europe may relate to enhancing the perception of a company's commitment to CSR. Such language may signal a strong commitment and the capability to implicit CSR agency, and enhance the effectiveness of CSR storytelling by managers. In addition, we highlight that the clout variable is strongly correlated with other pronouns (r = 0.898) (cf. supra Table 8), that may indicate that EU SR CEO letters tend to focus on consensual relationship-building more than their US counterparts, which is in line with our hypothesis (H2a). These results further provide evidence that the language in US SR CEO letters is more affective and emotional, which aligns with the distinct cultural characteristics of the explicit institutional framework and provides backing for our hypothesis (H2b). When controls are added in Model 7, Model 8 and Model 9, our results remain robust.

Table 11: Pairwise correlations

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
(1) Analytic	1.000																
(2) Clout	-0.461*	1.000															
(3) Tone	-0.285*	0.498*	1.000														
(4) SR	-0.343*	0.249*	0.146*	1.000													
(6) US_dum	-0.179*	0.290*	0.221*	0.000	1.000												
(7) ROA	-0.064*	0.084*	0.060*	-0.001	-0.009	1.000											
(8) ROE	-0.059*	0.035	0.012	0.000	0.047	0.334*	1.000										
(9) LOSS	0.057*	-0.058*	-0.045	0.001	-0.058*	-0.344*	-0.140*	1.000									
(10) LEV	-0.003	0.001	-0.029	-0.001	0.069*	-0.026	0.337*	0.078*	1.000								
(11) LIQ	0.017	0.080*	0.076*	0.000	0.123*	0.146*	-0.033	0.060*	-0.129*	1.000							
(12) SIZE	-0.073*	0.066*	-0.052*	0.000	0.175*	-0.219*	-0.108*	-0.031	0.023	-0.304*	1.000						
(13) TOBQ	-0.079*	0.123*	0.125*	0.000	0.047	0.733*	0.301*	-0.188*	-0.019	0.169*	-0.350*	1.000					
(14) ENVSOC	-0.033	-0.007	-0.042	0.000	-0.144*	0.047	-0.018	-0.073*	-0.070*	-0.195*	0.449*	0.002	1.000				
(15) ESG_contro	-0.097*	0.034	-0.060*	0.001	0.064*	-0.057*	-0.059*	-0.010	-0.043	-0.149*	0.523*	-0.105*	0.286*	1.000			
(16) Board_size	-0.010	0.014	-0.079*	0.000	0.027	-0.119*	-0.082*	-0.037	-0.001	-0.137*	0.474*	-0.180*	0.302*	0.284*	1.000		
(17) CEO_duality	-0.033	0.115*	0.101*	0.000	0.547*	0.004	-0.007	-0.074*	-0.010	-0.024	0.191*	0.049	-0.020	0.115*	0.186*	1.000	
(18) CSR_com	0.026	-0.041	-0.077*	0.000	-0.032	-0.007	0.027	-0.031	0.010	-0.107*	0.188*	-0.032	0.299*	0.104*	0.097*	-0.049	1.000

Correlations significant at p<0.01

Table 12: Regression results

VARIABLES	(1) Analytic	(2) Clout	(3) Tone	(4) Analytic	(5) Clout	(6) Tone	(7) Analytic	(8) Clout	(9) Tone
SR	-6.052***	5.748***	4.084***	-6.042***	7.320***	2.762***	-6.038***	7.318***	2.761**
	(0.318)	(0.413)	(0.534)	(0.437)	(0.566)	(0.734)	(0.434)	(0.564)	(0.729)
US_dum	-3.527***	6.924***	6.444***	-3.517***	8.590***	5.044***	-4.298***	9.084***	4.969***
	(0.353)	(0.459)	(0.594)	(0.475)	(0.616)	(0.798)	(0.522)	(0.679)	(0.877)
SR X US_dum				-0.020	-3.333***	2.801***	-0.024	-3.331***	2.801***
				(0.636)	(0.825)	(1.069)	(0.632)	(0.822)	(1.062)
ROA	-0.025	0.061	0.013	-0.025	0.061	0.0134	0.028	-0.031	-0.114*
	(0.030)	(0.039)	(0.050)	(0.030)	(0.039)	(0.0500)	(0.041)	(0.053)	(0.068)
LOSS	0.599	-0.717	-1.101	0.599	-0.721	-1.097	0.679	-0.839	-1.169
	(0.600)	(0.780)	(1.009)	(0.600)	(0.778)	(1.008)	(0.605)	(0.786)	(1.015)
LIQ	0.713**	0.450	0.362	0.713**	0.450	0.362	0.841***	0.273	0.219
	(0.312)	(0.405)	(0.524)	(0.312)	(0.404)	(0.524)	(0.314)	(0.408)	(0.527)
SIZE	-0.463***	0.811***	-0.614**	-0.463***	0.811***	-0.614**	-0.443***	1.123***	-0.149
	(0.143)	(0.186)	(0.241)	(0.143)	(0.186)	(0.241)	(0.172)	(0.224)	(0.290)
ENVSOC	0.003	-0.021	0.001	0.003	-0.021	0.010	0.000	-0.020	0.019
	(0.012)	(0.015)	(0.020)	(0.012)	(0.015)	(0.020)	(0.012)	(0.016)	(0.021)
ESG_contro							-1.090***	-0.517	-1.619**
							(0.390)	(0.507)	(0.655)
ROE							-0.509	-0.716	-0.967
							(0.480)	(0.624)	(0.806)
LEV							0.210**	-0.190	-0.333*
							(0.105)	(0.136)	(0.176)
TOBQ							-0.437*	1.025***	1.280***
							(0.261)	(0.339)	(0.438)

Board_size							0.044	-0.019	-0.260***
							(0.055)	(0.075)	(0.097)
CEO_duality							1.554***	-1.151**	-0.188
							(0.404)	(0.525)	(0.679)
CSR_com							1.249**	-1.377*	-2.753***
							(0.587)	(0.763)	(0.985)
Constant	98.73***	67.08***	82.44***	98.72***	66.29***	83.10***	96.86***	61.13***	71.35***
	(3.415)	(4.441)	(5.743)	(3.419)	(4.431)	(5.742)	(3.791)	(5.059)	(6.534)
Observations	2,472	2,472	2,472	2,472	2,472	2,472	2,472	2,472	2,472
R-squared	0.208	0.218	0.104	0.208	0.223	0.106	0.219	0.229	0.112

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

To further test the robustness of the region effect, we additionally run the models using region dummies. As an additional analysis, we use region dummies as *UK_dum* and *US_dum* and use *EU_dum* as the reference region. While our study did not initially formulate specific hypotheses regarding the language usage in UK CEO letters, due to the inconsistent classifications of the UK-style in the literature (e.g., Blindheim, 2012; Matten & Moon, 2008; Lee & Riffe, 2017) we further conduct an additional analysis. For example, Matten and Moon (2008) show that the UK still can be considered Europe, yet still highlight many differences in the institutional setting. According to Moon (2004), the demands of the UK stakeholders are shown to be more concerned on matters of sustainability and CSR than their European counterparts; hence ask for greater accountability and transparency. In addition, Blindheim (2012) argues that the extensive regulation in European countries leaves less space for voluntary actions compared to liberal market economies such as the UK or the US, which leads companies located in liberal market economies to report more explicitly about CSR. Drawing on the framework by Matten and Moon (2008), more recently, Lagore et al., (2020) show the unique position of the UK, which is in between explicit and implicit.

Table 13 shows the results with UK region controls. Our analysis reveals noteworthy findings regarding the effect of the UK interaction effect, distinguishing itself from the US, particularly for the rational-factual and personalized sense-giving language styles (Model 4 and Model 5). Model 4 shows that the UK institutional setting negatively moderates the use of rational-factual language, while the US institutional setting does not significantly affect the relationship. As Model 5 shows, the influence of the US and the UK institutional setting differs, particularly for the use of personalized sense-giving language (i.e., significant opposite directions), in comparison with the EU institutional setting. In our hypothesis, we expect that the implicit CSR setting (as in the EU) would constrain the use of personalized discourse and lead to more consensual relationship-building. In line with this, the more personalized discourse in the UK (which is a more explicit CSR business culture) in comparison with the EU provides partial support for our expectations. While this observation highlights the intriguing role of the UK institutional environment and provides evidence that the implicitness of the CSR environment encourages management to employ a more personalized sense-giving language, aligning with our initial expectations, the role of the US setting is counterinitiative. The negative moderation effect of the US institutional setting may relate to (possible) discourse about business models that integrate CSR initiatives which also help construct pragmatic legitimacy among shareholders. Finally, Model 6 also demonstrates significant interaction effects of both UK and US institutional settings on the use of assertive relationship-oriented language. In line with our expectations, our results show that European managers tend to use less assertive relationship-oriented language for storytelling, while in the UK and US storytelling mainly relies on such language. Overall, the separation of the UK observations enriches the results and provides a depth of analysis, revealing distinct language patterns that may relate to the use of different linguistic variables and rhetorical appeals, especially regarding the personalized, authoritative and confident language. Figure 2 further illustrates the interaction effects between disclosure genre (*SR*) and regions (*EU_dum*, *UK_dum* and *US_dum*) on rational-factual, personalized sense-giving and assertive relationship-oriented language styles (prediction lines).

Finally, we also use the OECD Environmental Policy Stringency Index as a control variable to account for variations in the institutional environment (untabulated). The index measures the extent to which environmental policies impose costs on polluting or environmentally harmful behavior, with values ranging from 0 (non-stringent) to 6 (stringent) (OECD, 2024; Rosati & Faria, 2019). Based on the index for the year 2020, the UK exhibited the most stringent policies (3.61), followed by the EU average (3.34) and the US (3.03). We conduct our analysis exclusively on SR, incorporating the control variable of the OECD

Environmental Policy Stringency Index for countries. The results indicate that while environmental stringency per does not significantly impact the use of rational-factual language (b = -0.183, p > 0.1); it positively influences the use of personalized sense-giving (b = 0.296, p < 0.05); and negatively influences the use of assertive relationship-oriented language (b = -0.455, p < 0.01). The results are somewhat consistent with the observations regarding the UK and US institutional settings: The UK with its more explicit CSR culture and stringent environmental policies tends to favor a personalized sense-giving approach in communication. Conversely, the US institutional effect, characterized by a pragmatic approach to CSR, may indicate a strategic shift towards more assertive and relationship-oriented style in response to lower stringent environmental regulations.

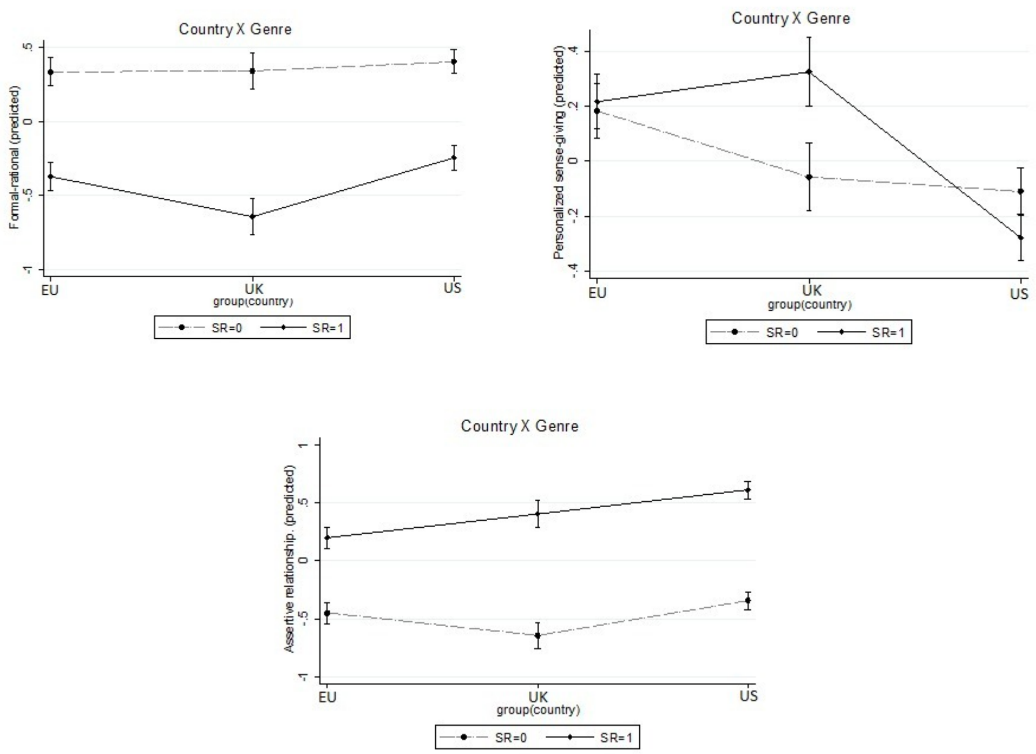

Figure 2: The effect of EU_dum, UK_dum and US_dum and SR (and FR) on the use of predicted appeals.
Both lines serve the purpose of highlighting the relative differences based on the binary categories (US_dum and UK_dum).

Table 13: Regression results

VARIABLES	(1) Ration. fact.	(2) Personal. sense-giv.	(3) Assert. relat.	(4) Ration. fact.	(5) Personal. sense-giv.	(6) Assert. relat.	(7) Ration. fact.	(8) Personal. sense-giv.	(9) Assert. relat.
SR	-0.741***	0.013	0.880***	-0.710***	0.033	0.651***	-0.710***	0.033	0.651***
	(0.037)	(0.039)	(0.035)	(0.066)	(0.068)	(0.065)	(0.066)	(0.069)	(0.062)
UK_dum	-0.160***	-0.041	-0.014	-0.023	-0.216***	-0.215***	0.005	-0.240***	-0.195***
	(0.053)	(0.056)	(0.051)	(0.074)	(0.078)	(0.071)	(0.076)	(0.080)	(0.072)
US_dum	0.089	-0.462***	0.223***	0.059	-0.362***	0.070	0.068	-0.294***	0.106
	(0.459)	(0.048)	(0.043)	(0.062)	(0.066)	(0.059)	(0.068)	(0.071)	(0.064)
SRxUK_dum				-0.275***	0.350***	0.400***	-0.276***	0.349***	0.399***
				(0.103)	(0.108)	(0.099)	(0.104)	(0.109)	(0.099)
SRxUS_dum				0.058	-0.202**	0.306***	0.059	-0.202**	0.306***
				(0.085)	(0.089)	(0.081)	(0.084)	(0.089)	(0.081)
ROA	-0.001	-0.008**	-0.007**	-0.001	-0.008*	-0.007**	0.009*	-0.004	-0.006
	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.005)	(0.005)	(0.005)
LOSS	0.031	0.036	-0.098	0.031	0.035	-0.098	0.059	0.061	-0.088
	(0.070)	(0.074)	(0.067)	(0.070)	(0.073)	(0.067)	(0.071)	(0.074)	(0.067)
LIQ	-0.041	0.027	0.069**	-0.041	0.028	0.069**	-0.051	0.018	0.053
	(0.036)	(0.038)	(0.035)	(0.036)	(0.038)	(0.035)	(0.037)	(0.039)	(0.035)
SIZE	0.024	0.033*	-0.019	0.024	0.033*	-0.019	0.006	0.009	-0.012
	(0.017)	(0.018)	(0.016)	(0.017)	(0.018)	(0.016)	(0.020)	(0.021)	(0.019)
ENVSOC	0.001	0.002	0.002*	0.001	0.002	0.002*	0.001	0.002	0.003**
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
ESG_contro							0.101**	0.098**	0.000

							(0.046)	(0.048)	(0.043)
ROE							-0.034	0.155***	-0.068
							(0.056)	(0.059)	(0.053)
LEV							-0.026**	-0.031**	-0.026**
							(0.012)	(0.013)	(0.012)
TOBQ							-0.083***	-0.049	-0.005
							(0.030)	(0.032)	(0.029)
Board_size							-0.001	0.006	-0.004
							(0.007)	(0.007)	(0.006)
CEO_duality							0.056	-0.124**	-0.034
							(0.048)	(0.049)	(0.045)
CSR_com							-0.035	-0.016	-0.118*
							(0.069)	(0.072)	(0.065)
Industry dummies	Included	Included	Included	Included	Included	Included	Included	Included	Included
Year dummies	Included	Included	Included	Included	Included	Included	Included	Included	Included
Constant	-0.061	-0.467	-0.298	-0.076	-0.478	-0.184	0.625	-0.065	-0.175
	(0.400)	(0.421)	(0.381)	(0.401)	(0.420)	(0.380)	(0.456)	(0.478)	(0.434)
Observations	2,472	2,472	2,472	2,472	2,472	2,472	2,472	2,472	2,472
R-squared	0.159	0.069	0.238	0.162	0.080	0.244	0.170	0.088	0.249

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

5. Discussion and conclusion

The purpose of the current study is to examine top management's use of language in FR and SR CEO letters. Language is viewed as a means through which power relations are expressed (Brennan & Merkl-Davies, 2014). Studying language in different corporate (sub)genres presents insights into management's views on different contents and the way they impose their thoughts (Mäkelä & Laine, 2011). Prior literature documents that FR and SR are distinct corporate genres; hence, the language and IM expressions differ substantially between them (Caliskan et al., 2021). Research on the language used in the mentioned subgenres, however, typically exhibits certain limitations such as small sample sizes, often adopting methods like case studies or manual content analysis (e.g., Blanc et al., 2019; Caliskan et al., 2021), a narrow selection of linguistic attributes, such as stance markers or readability (e.g., Barkemeyer et al., 2014; Fuoli, 2008), a concentration on specific sectors, such as metal and mining (e.g., Mäkelä & Laine, 2011), and limited geographic regions, such as Finland and Turkey (e.g., Caliskan et al., 2021; Mäkelä & Laine, 2011). Consequently, prior research lacks a systematic understanding of the motives behind the linguistic differences in FR and SR. Unlike previous studies that compare (sub)genres, our paper offers a more comprehensive perspective for understanding linguistic distinctions employing a larger and more diverse sample.

The findings of this study make several contributions to the current literature and enhance the overall understanding of corporate communication strategies. First and from a theoretical perspective, our contribution to the literature is rooted in legitimacy theory. We posit that the concept of legitimacy varies in accordance with shared conventions and expectations among stakeholders. Legitimacy plays a central role in business communication (Deegan, 2002; Martins et al., 2019; Suchman, 1995). Nevertheless, it is important to recognize that legitimacy encompasses various forms (Suchman, 1995). Employing the generic term “legitimacy” without discerning these distinctions would result in the oversight of crucial nuances related to linguistic variations across different genres. In our study, we argue that linguistic differences between FR and SR CEO letters mainly relate to shared conventions and expectations of distinct targeted audiences and therefore, different legitimacy concerns, regulatory requirements and nature of information.

Second, our findings, derived from factor analysis, offer empirical evidence regarding the significance of specific linguistic variables within distinct subgenres. In line with our expectations, FR CEO letters reflect a more rational and logical stance in language use. The rational-factual appeal loads highly on numbers, quantities, time, conjunctions and differentiation. This demonstrates that FR CEO letters are more factual and time-based, and depend on descriptive and logical thinking. The concrete outcomes and impersonal speech are expected to satisfy the pragmatic needs of the FR audience, such as shareholders and investors. This indicates that top management definitely focuses on maintaining or gaining pragmatic legitimacy among the FR audience, who also have knowledge and expertise to interpret financial accounting performance explanations. The use of such descriptive language in FR CEO letters may indicate that FR represents a more functional accountability mechanism compared to SR, which is also due to its nature of information. Yet, we also highlight the presence of affective language in our rational-factual factor, which may also indicate that rational-factual does not necessarily mean the absence of self-serving IM in FR CEO letters. The general content revolves around principles associated with general accounting and financial reasoning, providing information on financial indicators, such as growth, profitability and operational performance.

For SR CEO letters, we document that they reflect a more informal and conversational storytelling style compared to FR CEO letters. Information on a company’s SP and commitment to social, environmental and ethical issues is the main focus of SR (Barkemeyer et al., 2014; Mäkelä & Laine, 2011). This kind of information typically concerns a broad range of stakeholders including civil society and NGOs due to the increasing sustainability awareness. This broad range of stakeholders is not homogenously distributed in terms of their knowledge, education and expertise (Barkemeyer et al., 2014; Fuoli, 2018; Lindgren et al., 2021; Mäkelä & Laine, 2011). As our results show, the diversity in the target audience should be constraining top management to use technical language with jargon to build up or maintain legitimacy. Therefore, the shared conventions and expectations associated with SR CEO letters provide greater flexibility for opportunistic storytelling behavior, resembling more overt symbolic IM. Organizations often seek moral legitimacy among a broad range of stakeholders and are expected to show they are *doing the right thing*. For that, top management tries to present the company as a good corporate citizen and a credible environmental steward, mainly through a discourse of goodwill (Lin-Hi & Blumberg, 2016; Bhatia, 2012). In line with our expectations, we argue that to sustain or gain moral legitimacy, top management tends to use emotional and personal arguments for robust and effective communication with stakeholders and call for values such as humanism, benevolence and diversity in SR. Besides moral legitimacy concerns, we recognize that the absence of established standards in SR, as well as the ambiguous character of SP (Boiral et al., 2020; Caliskan et al., 2021) also play a role and to some extent compel the adoption of informal and conversational storytelling language choice in SR narratives.

Third, as the literature shows, storytelling can be constructed in various ways due to strategic reasons in corporate narratives (Martins et al., 2019; Spear & Roper, 2013), which is expected to vary according to regional-level institutional settings. To investigate the potential variations in storytelling within SR CEO letters, we apply the implicit-explicit framework developed by Matten and Moon (2008). Our factor analysis demonstrates two informal conversational storytelling appeals (i.e., personalized sense-giving and assertive relationship-oriented) and results show that both of them tend to be more prominent in SR, supporting our expectations. The personalized sense-giving appeal indicates a more personalized, honest and trustworthy discourse, which would help top management gain moral legitimacy from a broad range of stakeholders. This stance relies on the factor that highly loads on cognitive processes indicators (i.e., discrepancy, tentative, differentiation and certitude), which help managers to enhance their authoritative image while showing a willingness to honestly address challenging aspects of the issue (e.g., sustainability or CSR policies of companies), and "*I*" pronouns that indicate CEO's personal commitments with neutral wording. Our results indicate that this storytelling approach is more pronounced in EU CEO letters compared to the US CEO letters. This observation may relate to the presence of implicit checks and balances within European business culture and government policies would lead us to anticipate a greater prominence of this storytelling style in EU CEO letters, where top management needs to explain the way they integrate CSR to their business models. This observation may suggest a potential alignment between European SR policies and the managerial responses aimed at meeting these demands. Additionally, this may indicate that the EU SR policies promote transparency and accountability. However, it is important to note that our results do not necessarily demonstrate the adoption of these policies or confirm that EU SR CEO letters consistently fulfill the intended purpose of SR policies, as managers tend to use cognitive shading and maintain -to some extent- cautious stance in their communications, as well. In addition, our main analysis alone fails to directly test our hypothesis that relates to relationship-building alone. We further elaborate on this in the additional analysis section using the LIWC summary variables. In line with our expectations, LIWC results show that *clout* (that is highly correlated with "other pronouns") and therefore relationship orientation is higher in the EU SR CEO letters in comparison with the US SR CEO letters. Further, our results also demonstrate that UK-style reporting may not always be in the middle of the implicit-explicit continuum; hence, based on what is being investigated, UK observations must be carefully considered where to cluster. We show that top management in the UK tends to use significantly higher personalized language in comparison with top management from the EU. While this may relate to the business models and pragmatic legitimacy expectations in the UK, our results also show that the personalized language is negatively moderated by the US institutional effect. This is counterintuitive, and further studies on the differences in personalized language between the UK and the US would be worthwhile. The regional institutional setting presents another significant consideration, concerning potential environmental litigation risks in the near future. Currently, in the US, SR is not mandatory but, in the EU, the new legislation (CSRD - 2022/2464/EU) mandates large and listed companies to disclose SR being effective from 2025 covering the operations in the fiscal year 2024. Subsequent research may investigate how these regulations could influence IM within SR practices.

The informal conversational storytelling appeal that our factor analysis demonstrates relates to high loadings on the use of ‘other pronouns’ and affective language. This factor combines both solidarity, due to ‘other pronouns’, and emotions, due to insight and affective words. In addition, high loadings of conjunctions and causality statements may indicate build-up and longer stories. Mainly due to these two variables, this style tends to proxy an assertive and relationship-oriented stance. We argue that this style is more prominent in the US SR CEO letters. According to Matten and Moon (2008), US-style CSR is explicit, where companies are encouraged to integrate CSR concerns and strategies into their business activities voluntarily. Voluntary and strategic CSR activities lead to a greater space for overall self-serving IM, where assertive acclaiming and overly positive tone management would play a role, in the absence of implicit checks and balances. Our results support our expectations and evoke further discussions about the effectiveness of purely voluntary CSR strategies and communication, as in the US-style. Our results provide empirical evidence, in that sense, that government regulations and stakeholder scrutiny in different regions shape SR communication, and possibly activities, as well. Although we have not made specific predictions in our hypotheses, our results also demonstrate the impact of regional-level institutional settings on the utilization of appeals in FR (when SR = 0).

Although we abstain from formulating specific hypotheses regarding storytelling behavior in FR CEO letters, our results, as illustrated in Figure 1 and Figure 2, yield insights into the nature of storytelling within this context, as well. Overall, the empirical findings presented in this study highlight different communication strategies employed by top management in FR and SR CEO letters. This difference is attributed to inherent differences in genre characteristics, that reflect stakeholder expectations, and institutional setting, that reflect regional-level regulations on the use of disclosures' narrative sections. The direct comparative analysis of CEO letters across diverse genres enables a nuanced understanding of their accountability and transparency related features. Our exploration of genre differences leads us to argue that SR, as a genre, continues to face scrutiny regarding its informational quality. Furthermore, our empirical results indicate that regional-level institutional settings might have a potential influence on the utilization of SR as an accountability mechanism. While our research does not directly provide insights into the application of IM within these subgenres, this study provides key insights into possible self-serving rhetorical IM practices. In summary, our study offers practical insights for policymakers, investors and analysts by providing a nuanced understanding of linguistic differences due to genre and regional-level differences, anticipating the impact of upcoming regulations, such as mandated SR as of 2025, and offering indicators regarding strategic investment and risk assessment decisions through understanding distinct storytelling practices.

The main limitation of our research is mainly related to the availability of SR CEO letters and the lack of trustable data sources that have the collection of multiple companies.

While we tried to reach companies' disclosures manually on different web pages and company internet sites, we still have a major miss of FR and SR CEO letters. We also acknowledge that we do not control for individual-level CEO motivations, which may affect CSR disclosure behaviors (Lassoued & Khanchel, 2022), as well as control variables related to demand for information (e.g., analyst following, ownership structure) due to lack of data. In addition, further studies could explore additional dimensions of policy stringency per region and country. Finally, the dictionary-based textual analysis approach has certain limitations, such as overlooking the real meaning and content. While factor analysis may overcome this limitation to a certain extent, further research may overcome this limitation by integrating style and content variables in the analyses to understand the multidimensional features of sustainability-related information. The application of machine learning and AI techniques holds the potential to provide solutions for advanced text analysis in this regard. Theoretically, a greater focus on types of legitimacy could produce interesting findings in future studies. The insights into significant linguistic variables for different subgenres, and especially for SR, may be explored further in the future. In addition, we highlight that considerably more work will be needed to determine to what extent non-quantification would be considered as IM in SR. Practically, this research is useful for regulatory authorities to make corporate genres work as more functioning accountability mechanisms. Insights gained from our study about linguistic differences between FR and SR can aid in complying with regulatory reporting requirements for transparent and credible communication in corporate disclosures. With insights into how regional-level institutional settings influence communication strategies, companies can anticipate potential risks related to environmental litigation and changing regulations. The users of both FR and SR would also improve their evaluation criteria through the lenses we provided. Overall, our study encourages a more accountable and ethical approach to communication.

Reference List

Adaji, I., & Olakanmi, O. (2019). *Evolution of Emotions and Sentiments in an Online Learning Community* Proceedings of the first International Workshop on Supporting Lifelong Learning co-located with the 20th International Conference on Artificial Intelligence (AIED 2019), Chicago, USA.

Aerts, W. (2005). Picking up the pieces: impression management in the retrospective attributional framing of accounting outcomes. *Accounting, Organizations and Society, 30*(6), 493-517. https://doi.org/10.1016/j.aos.2004.07.001

Aerts, W., & Cormier, D. (2009). Media legitimacy and corporate environmental communication. *Accounting, Organizations and Society, 34*(1), 1-27. https://doi.org/10.1016/j.aos.2008.02.005

Aerts, W., & Yan, B. (2017). Rhetorical impression management in the letter to shareholders and institutional setting. *Accounting, Auditing & Accountability Journal, 30*(2), 404-432. https://doi.org/10.1108/aaaj-01-2015-1916

Aguilera, R. V., Williams, C. A., Conley, J. M., & Rupp, D. E. (2006) Corporate governance and social responsibility: A comparative analysis of the U.K. and the U.S. *Corporate governance: An International Review,* 14(3), 147-157. https://doi.org/10.1111/j.1467-8683.2006.00495.x

Al Hawaj, A. Y., & Buallay, A. M. (2021). A worldwide sectorial analysis of sustainability reporting and its impact on firm performance. *Journal of Sustainable Finance & Investment, 12*(1), 62-86. https://doi.org/10.1080/20430795.2021.1903792

Amernic, J. H., & Craig, R. J. (2007). Guidelines for CEO-speak: editing the language of corporate leadership. *Strategy & Leadership, 35*(3), 25-31. https://doi.org/10.1108/10878570710745802

Aouadi, A., & Marsat, S. (2018). Do ESG Controversies Matter for Firm Value? Evidence from International Data. *Journal of Business Ethics, 151*(4), 1027-1047. https://doi.org/10.1007/s10551-016-3213-8

Araujo, T., & Kollat, J. (2018). Communicating effectively about CSR on Twitter. *Internet Research, 28*(2), 419-431. https://doi.org/10.1108/IntR-04-2017-0172

Băndoi, A., Bocean, C. G., Del Baldo, M., Mandache, L., Mănescu, L. G., & Sitnikov, C. S. (2021). Including Sustainable Reporting Practices in Corporate Management Reports: Assessing the Impact of Transparency on Economic Performance. *Sustainability, 13*(2). https://doi.org/10.3390/su13020940

Barkemeyer, R., Comyns, B., Figge, F., & Napolitano, G. (2014). CEO statements in sustainability reports: Substantive information or background noise? *Accounting Forum, 38*(4), 241-257. https://doi.org/10.1016/j.accfor.2014.07.002

Beattie, V., & Jones, M. (1997). A comparative study of the use of financial graphs in the corporate annual reports of major U.S. and U.K. companies. *Journal of International Financial Management & Accounting,* 8(1), 33-68. https://doi.org/10.1111/1467-646X.00016

Bhatia, A. (2012). The Corporate Social Responsibility Report: The Hybridization of a "Confused" Genre (2007-2011). *IEEE Transaction on Professional Communication,* 55(3). https://doi.org/10.1109/TPC.2012.2205732

Bickel, M. W. (2017). A new approach to semantic sustainability assessment: text mining via network analysis revealing transition patterns in German municipal climate action plans. *Energy, Sustainability and Society, 7*(1). https://doi.org/10.1186/s13705-017-0125-0

Birkey, R. N., Michelon, G., Patten, D. M., & Sankara, J. (2016). Does assurance on CSR reporting enhance environmental reputation? An examination in the U.S. context. *Accounting Forum, 40*(3), 143-152. https://doi.org/10.1016/j.accfor.2016.07.001

Blanc, R., Cho, C. H., Sopt, J., & Branco, M. C. (2019). Disclosure Responses to a Corruption Scandal: The Case of Siemens AG. *Journal of Business Ethics, 156*(2), 545-561. https://doi.org/10.1007/s10551-017-3602-7

Blindheim, B.-T. (2012). Institutional Models of Corporate Social Responsibility. *Business & Society, 54*(1), 52-88. https://doi.org/10.1177/0007650312443961

Boiral, O. (2016). Accounting for the Unaccountable: Biodiversity Reporting and Impression Management. *Journal of Business Ethics, 135*(4), 751-768. https://doi.org/10.1007/s10551-014-2497-9

Boiral, O., Brotherton, M.-C., & Talbot, D. (2020). Building trust in the fabric of sustainability ratings: An impression management perspective. *Journal of Cleaner Production, 260*. https://doi.org/10.1016/j.jclepro.2020.120942

Boyd, R. L. (2017). Psychological Text Analysis in the Digital Humanities. In *Data Analytics in Digital Humanities* (pp. 161-189). https://doi.org/10.1007/978-3-319-54499-1_7

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). *The Development and Psychmetric Propoerties of LIWC-22*. The University of Texas at Austin.

Braam, G., & Peeters, R. (2018). Corporate Sustainability Performance and Assurance on Sustainability Reports: Diffusion of Accounting Practices in the Realm of Sustainable Development. *Corporate Social Responsibility and Environmental Management, 25*(2), 164-181. https://doi.org/10.1002/csr.1447

Brennan, N. M., & Merkl-Davies, D. M. (2014). Rhetoric and argument in social and environmental reporting: the Dirty Laundry case. *Accounting, Auditing & Accountability Journal, 27*(4), 602-633. https://doi.org/10.1108/aaaj-04-2013-1333.

Breton, G., & Taffler, R. J. (2001). Accounting information and analyst stock recommendation decisions: a content analysis approach. *Accounting and Business Research, 31*(2), 91-101. https://doi.org/10.1080/00014788.2001.9729604

Buallay, A. (2019). Between cost and value: Investigating the effects of sustainability reporting on a firm's performance. *Journal of Applied Accounting Research, 20*(4), 481-496. https://doi.org/10.1108/jaar-12-2017-0137

Caliskan, A. O., Esen, E., & Barkemeyer, R. (2021). Impression management tactics in the CEO statements of Turkish sustainability reports. *Business Ethics, the Environment & Responsibility, 30*(4), 485-506. https://doi.org/10.1111/beer.12374

Chelli, M., Durocher, S., & Fortin, A. (2018). Normativity in Environmental Reporting: A Comparison of Three Regimes. *Journal of Business Ethics, 149*(2), 285-311. https://doi.org/10.1007/s10551-016-3128-4

Cho, C. H., Guidry, R. P., Hageman, A. M., & Patten, D. M. (2012). Do actions speak louder than words? An empirical investigation of corporate environmental reputation. *Accounting, Organizations and Society, 37*(1), 14-25. https://doi.org/10.1016/j.aos.2011.12.001

Cho, C. H., Michelon, G., Patten, D. M., & Roberts, R. W. (2015). CSR disclosure: the more things change…? *Accounting, Auditing & Accountability Journal, 28*(1), 14-35. https://doi.org/10.1108/aaaj-12-2013-1549

Cho, C. H., & Patten, D. M. (2007). The role of environmental disclosures as tools of legitimacy: A research note. *Accounting, Organizations and Society, 32*(7-8), 639-647. https://doi.org/10.1016/j.aos.2006.09.009

Cho, C. H., & Roberts, R. W. (2010). Environmental reporting on the internet by America's Toxic 100: Legitimacy and self-presentation. *International Journal of Accounting Information Systems, 11*(1), 1-16. https://doi.org/10.1016/j.accinf.2009.12.003

Chung, C. K., & Pennebaker, J. W. (2007). The Psychological Functions of Function Words. In K. Fiedler (Ed.), *Social Communication* (pp. 343-359). Psychology Press.

Clarkson, P. M., Li, Y., Richardson, G. D., & Vasvari, F. P. (2008). Revisiting the relation between environmental performance and environmental disclosure: An empirical analysis. *Accounting, Organizations and Society, 33*(4-5), 303-327. https://doi.org/10.1016/j.aos.2007.05.003

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2014). Linguistic markers of psychological change surrounding. *Psychological Science, 15*(10), 687-693. https://doi.org/10.1111/J.0956-7976.2004.00741.X

Cong, Y., Freedman, M., & Park, J. D. (2014). Tone at the top: CEO environmental rhetoric and environmental performance. *Advances in Accounting, 30*(2), 322-327. https://doi.org/10.1016/j.adiac.2014.09.007

Cormier, D., Aerts, W., Ledoux, M.-J., & Magnan, M. (2009). Attributes of social and human capital disclosure and information asymmetry between managers and investors. *Canadian Journal of Administrative Sciences / Revue Canadienne des Sciences de l'Administration, 26*(1), 71-88. https://doi.org/10.1002/cjas.89

Cormier, D., & Magnan, M. (2003). Environmental reporting management: a continental European perspective. *Journal of Accounting and Public Policy, 22*(1), 43-62. https://doi.org/10.1016/s0278-4254(02)00085-6

DasGupta, R. (2021). Financial performance shortfall, ESG controversies, and ESG performance: Evidence from firms around the world. *Finance Research Letters*. https://doi.org/10.1016/j.frl.2021.102487

Davies, M. (2011). Concept mapping, mind mapping and argument mapping: what are the differences and do they matter? *Higher Education, 62*(3), 279-301. https://doi.org/10.1007/s10734-010-9387-6

De Villiers, C., & Maroun, W. (2018). *Sustainability accounting and integrated reporting*. Routledge.

Deegan, C. (2002). The legitimising effect of social and environmental disclosures – a theoretical foundation. *Accounting, Auditing & Accountability Journal, 15*(3), 282-311. https://doi.org/10.1108/09513570210435852

Del Giudice, A., & Rigamonti, S. (2020). Does Audit Improve the Quality of ESG Scores? Evidence from Corporate Misconduct. *Sustainability, 12*(14). https://doi.org/10.3390/su12145670

Directive (EU) 2022/2464 of the European Parliament and of the Council of 14 December 2022 amending Regulation (EU) No 537/2014, Directive 2004/109/EC, Directive 2006/43/EC and Directive 2013/34/EU, as regards corporate sustainability reporting. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022L2464.

Dorfleitner, G., Kreuzer, C., & Sparrer, C. (2020). ESG controversies and controversial ESG: about silent saints and small sinners. *Journal of Asset Management, 21*(5), 393-412. https://doi.org/10.1057/s41260-020-00178-x

Du, S., & Yu, K. (2020). Do Corporate Social Responsibility Reports Convey Value Relevant Information? Evidence from Report Readability and Tone. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-020-04496-3

Duncan, S. Y., Chohan, R., & Ferreira, J. J. (2019). What makes the difference? Employee social media brand engagement. *Journal of Business & Industrial Marketing, 34*(7), 1459-1467. https://doi.org/10.1108/jbim-09-2018-0279

Dunne, N. J., Brennan, N. M., & Kirwan, C. E. (2021). Impression management and Big Four auditors: Scrutiny at a public inquiry. *Accounting, Organizations and Society, 88*. https://doi.org/10.1016/j.aos.2020.101170

Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics, 64*(2-3), 221-245. https://doi.org/10.1016/j.jacceco.2017.07.002

Edgar, V. C., Beck, M., & Brennan, N. M. (2018). Impression management in annual report narratives: the case of the UK private finance initiative. *Accounting, Auditing & Accountability Journal, 31*(6), 1566-1592. https://doi.org/10.1108/aaaj-10-2016-2733

Fehre, K., & Weber, F. (2016). Challenging corporate commitment to CSR. *Management Research Review, 39*(11), 1410-1430. https://doi.org/10.1108/mrr-03-2015-0063

Ferri, L. M. (2017). The influence of the institutional context on sustainability reporting. A cross-national analysis. *Social Responsibility Journal, 13*(1), 24-47. https://doi.org/10.1108/srj-11-2015-0172

Fiandrino, S., & Tonelli, A. (2021). A Text-Mining Analysis on the Review of the Non-Financial Reporting Directive: Bringing Value Creation for Stakeholders into Accounting. *Sustainability, 13*(2). https://doi.org/10.3390/su13020763

Fifka, M. S., & Drabble, M. (2012). Focus and Standardization of Sustainability Reporting – A Comparative Study of the United Kingdom and Finland. *Business Strategy and the Environment, 21*, 455-474. https://doi.org/10.1002/bse.1730

François, A., Bayle, E., & Gond, J.-P. (2018). A multilevel analysis of implicit and explicit CSR in French and UK professional sport. *European Sport Management Quarterly, 19*(1), 15-37. https://doi.org/10.1080/16184742.2018.1518468

Fuoli, M. (2018). Building a Trustworthy Corporate Identity: A Corpus-Based Analysis of Stance in Annual and Corporate Social Responsibility Reports. *Applied Linguistics, 39*(6), 846-885. https://doi.org/10.1093/applin/amw058

García-Meca, E., & Martínez-Ferrero, J. (2021). Is SDG reporting substantial or symbolic? An examination of controversial and environmentally sensitive industries. *Journal of Cleaner Production, 298*. https://doi.org/10.1016/j.jclepro.2021.126781

García-Sánchez, I. M., & Araújo-Bernardo, C. A. (2019). What colour is the corporate social responsibility report? Structural visual rhetoric, impression management strategies, and stakeholder engagement. *Corporate Social Responsibility and Environmental Management, 27*(2), 1117-1142. https://doi.org/10.1002/csr.1869

Garzone, G. (2005). Letters to shareholders and chairman's statements: Textual variability and generic integrity. In G. P. & M. Gotti (Eds.), *Genre Variation in Business Letters* (pp. 179-204). Peter Lang.

Hahn, R., & Lülfs, R. (2014). Legitimizing Negative Aspects in GRI-Oriented Sustainability Reporting: A Qualitative Analysis of Corporate Disclosure Strategies. *Journal of Business Ethics, 123*(3), 401-420. https://doi.org/10.1007/s10551-013-1801-4

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate Data Analysis, Englewood Cliffs, NJ: Prentice Hall.

Harjoto, M. A., Jadallah, J., Laksmana, I., & Lee, W. E. (2021). Corporate Social Responsibility Reporting: Does Writing Style Matter? *International Journal of Accounting & Finance Review, 6*(1), 34-40.

Hawn, O., & Ioannou, I. (2016). Mind the gap: The interplay between external and internal actions in the case of corporate social responsibility. *Strategic Management Journal, 37*(13), 2569-2588. https://doi.org/10.1002/smj.2464

Hossain, M., Islam, M. T., Momin, M. A., Nahar, S., & Alam, M. S. (2019). Understanding Communication of Sustainability Reporting: Application of Symbolic Convergence Theory (SCT). *Journal of Business Ethics, 160*(2), 563-586. https://doi.org/10.1007/s10551-018-3874-6

Hotho, J. J. (2014). From Typology to Taxonomy: A Configurational Analysis of National Business Systems and their Explanatory Power. *Organization Studies, 35*(5), 671-702. https://doi.org/10.1177/0170840613502767

Hummel, K. (2017). *Corporate Social Responsibility Disclosure in Europe and the United States: Investigating the Implicit-Explicit-Framework Using Textual Analysis* 40th European Accounting Association (EAA) Annual Congress 2017, Valencia, Spain.

Hummel, K., & Rötzel, P. (2019). Mandating the Sustainability Disclosure in Annual Reports—Evidence from the United Kingdom. *Schmalenbach Business Review, 71*(2), 205-247. https://doi.org/10.1007/s41464-019-00069-8

Hyland, K. (1998). Exploring Corporate Rhetoric: Metadiscourse in the CEO's Letter. *The Journal of Business Communication, 35*(2), 224245. https://doi.org/10.1177/002194369803500203

Hyland, K. (2005). *Metadiscourse : exploring interaction in writing*. Continuum.

Ioannou, I., & Serafeim, G. (2012). What drives corporate social performance? The role of nation-level institutions. *Journal of International Business Studies, 43*(9), 834-864. https://doi.org/10.1057/jibs.2012.26

Jackson, G., & Apostolakou, A. (2010). Corporate Social Responsibility in Western Europe: An Institutional Mirror or Substitute? *Journal of Business Ethics, 94*(3), 371-394. https://doi.org/10.1007/s10551-009-0269-8

Järvinen, J. T., Laine, M., Hyvönen, T., & Kantola, H. (2020). Just Look at the Numbers: A Case Study on Quantification in Corporate Environmental Disclosures. *Journal of Business Ethics, 175*(1), 23-44. https://doi.org/10.1007/s10551-020-04600-7

Jonäll, K., & Rimmel, G. (2010). CEO letters as legitimacy builders: coupling text to numbers. *Journal of Human Resource Costing & Accounting, 14*(4), 307-328. https://doi.org/10.1108/14013381011105975

Jordan, K. N., Sterling, J., Pennebaker, J. W., & Boyd, R. L. (2019, Feb 26). Examining long-term trends in politics and culture through language of political leaders and cultural institutions. *Proc Natl Acad Sci U S A, 116*(9), 3476-3481. https://doi.org/10.1073/pnas.1811987116

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology, 33*(2), 125-143. https://doi.org/10.1177/0261927x13502654

LaGore, W., Mahoney, L., & Thorne, L. (2020). An implicit-explicit examination of differences in CSR practices between the USA and Europe. *Society and Business Review, 15*(3), 165-187. https://doi.org/10.1108/sbr-10-2019-0129

Lassoued, N., & Khanchel, I. (2022). Voluntary CSR disclosure and CEO narcissism: the moderating role of CEO duality and board gender diversity. *Review of Managerial Science, 17*(3), 1075-1123. https://doi.org/10.1007/s11846-022-00555-3

Lee, T. H., & Riffe, D. (2019). Business News Framing of Corporate Social Responsibility in the United States and the United Kingdom: Insights From the Implicit and Explicit CSR Framework. *Business & Society, 58*(4), 683-711. https://doi.org/1.o0r.g1/107.171/0770/0700675605301371766996 6314

Leibbrand, M. P. (2015). The language of executive financial discourse. *Studies in Communication Sciences, 15*(1), 45-52. https://doi.org/10.1016/j.scoms.2015.03.006

Lin-Hi, N., & Blumberg, I. (2016). The Link Between (Not) Practicing CSR and Corporate Reputation: Psychological Foundations and Managerial Implications. *Journal of Business Ethics, 150*(1), 185-198. https://doi.org/10.1007/s10551-016-3164-0

Lindgren, C., Huq, A. M., & Carling, K. (2021). Who Are the Intended Users of CSR Reports? Insights from a Data-Driven Approach. *Sustainability, 13*(3). https://doi.org/10.3390/su13031070

Liu, P., & Nguyen, H. T. (2020). CEO characteristics and tone at the top inconsistency. *Journal of Economics and Business, 108*. https://doi.org/10.1016/j.jeconbus.2019.105887

Mäkelä, H., & Laine, M. (2011). A CEO with many messages: Comparing the ideological representations provided by different corporate reports. *Accounting Forum, 35*(4), 217-231. https://doi.org/10.1016/j.accfor.2011.06.008

Marais, M. (2012). CEO rhetorical strategies for corporate social responsibility (CSR). *Society and Business Review, 7*(3), 223-243. https://doi.org/10.1108/17465681211271314

Mardini, G. H., & Lahyani, F. E. (2020). Impact of firm performance and corporate governance mechanisms on intellectual capital disclosures in CEO statements. *Journal of Intellectual Capital, 23*(2), 290-312. https://doi.org/10.1108/jic-02-2020-0053

Martins, A., Gomes, D., & Branco, M. C. (2021). Managing Corporate Social and Environmental Disclosure: An Accountability vs. Impression Management Framework. *Sustainability, 13*(1). https://doi.org/10.3390/su13010296

Martins, A., Gomes, D., Oliviera, L., & Ribeiro, J. L. (2019). Telling a success story through the president's letter. *Qualitative Research in Accounting and Management,* 16(), 403-433. https://doi.org/10.1108/QRAM-03-2018-0018

Mason, M., & Mason, R. D. (2012). Communicating a Green Corporate Perspective. *Journal of Business and Technical Communication, 26*(4), 479-506. https://doi.org/10.1177/1050651912448872

Matten, D., & Moon, J. (2008). "Implicit" and "Explicit" CSR: A Conceptual Framework for a Comparative Understanding of Corporate Social Responsibility. *Academy of Management Journal, 33*(2), 404-424.

Merkl-Davies, D. M., & Brennan, N. M. (2007). Discretionary disclosure strategies in corporate narratives : incremental information or impression management? *Journal of Accounting Literature, 26*(116-196).

Merkl-Davies, D. M., Brennan, N. M., & McLeay, S. J. (2011). Impression management and retrospective sense-making in corporate narratives. *Accounting, Auditing & Accountability Journal, 24*(3), 315-344. https://doi.org/10.1108/09513571111124036

Michelon, G., Pilonato, S., & Ricceri, F. (2015). CSR reporting practices and the quality of disclosure: An empirical analysis. *Critical Perspectives on Accounting, 33*, 59-78. https://doi.org/10.1016/j.cpa.2014.10.003

Mittelbach-Hörmanseder, S., Hummel, K., & Rammerstorfer, M. (2020). The information content of corporate social responsibility disclosure in Europe: an institutional perspective. *European Accounting Review, 30*(2), 309-348. https://doi.org/10.1080/09638180.2020.1763818

Moon, J. (2004). *Government as a driver of corporate social responsibility: The UK in comparative perspective* International Centre for Corporate Social Responsibility (ICCSR), Nottingham, UK.

Moses, E., Che-Ahmad, A., Abdulmalik, S. O., & Ntim, C. G. (2020). Board governance mechanisms and sustainability reporting quality: A theoretical framework. *Cogent Business & Management, 7*(1). https://doi.org/10.1080/23311975.2020.1771075

Na, H. J., Lee, K. C., Choi, S. U., & Kim, S. T. (2020). Exploring CEO Messages in Sustainability Management Reports: Applying Sentiment Mining and Sustainability Balanced Scorecard Methods. *Sustainability, 12*(2). https://doi.org/10.3390/su12020590

OECD. (2024). Environmental Policy Stringency Index. https://stats.oecd.org/Index.aspx?DataSetCode=EPS. (accessed on: 21 March 2024).

Öge, K. (2016). Which transparency matters? Compliance with anti-corruption efforts in extractive industries. *Resources Policy, 49*, 41-50. https://doi.org/10.1016/j.resourpol.2016.04.001

Oliver, K. M., Houchins, J. K., Moore, R. L., & Wang, C. (2020). Informing Makerspace Outcomes Through a Linguistic Analysis of Written and Video-Recorded Project Assessments. *International Journal of Science and Mathematics Education, 19*(2), 333-354. https://doi.org/10.1007/s10763-020-10060-2

Orlitzki, M., & Benjamin, J. D. (2001). Corporate Social Performance and Firm Risk: A Meta-Analytic Review. *Business & Society, 40*(4), 369-396. https://doi.org/10.1177/000765030104000402

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*.

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: the case of college admissions essays. *PLoS One, 9*(12), e115844. https://doi.org/10.1371/journal.pone.0115844

Pennebaker, J. W., Mayne, T. J., & Francis, M. E. (1997). Linguistic Predictors of Adaptive Bereavement. *Journal of Personality and Social Psychology, 72*(4), 863-871. https://doi.org/10.1037/0022-3514.72.4.863

Perrini, F. (2006). The Practitioner's Perspective on Non-Financial Reporting. *California Management Review, 48*(2), 73-103.

Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. In *Communication and Persuasion*. Springer Series in Social Psychology. https://doi.org/10.1007/978-1-4612-4964-1_1

Pizzi, S., Del Baldo, M., Caputo, F., & Venturelli, A. (2021). Voluntary disclosure of Sustainable Development Goals in mandatory non-financial reports: The moderating role of cultural dimension. *Journal of International Financial Management & Accounting*, 33(1). https://doi.org/10.1111/jifm.12139

Puroila, J., & Mäkelä, H. (2019). Matter of opinion: Exploring the socio-political nature of materiality in disclosures in sustainability reporting. *Accounting, Auditing & Accountability Journal, 32*(4), 1043-1072. https://doi.org/10.1108/aaaj-11-2016-2788

Ribando, J. M., & Bonne, G. (2010). *A New Quality Factor: Finding Alpha with Asset4 ESG Data*.

Rosati, F., & Faria, L. G. D. (2019). Addressing the SDGs in sustainability reports: The relationship with institutional factors. *Journal of Cleaner Production, 215*, 1312-1326. https://doi.org/10.1016/j.jclepro.2018.12.107

Săndulescu, M.-S. (2021). Sustainability reporting and impression management: A case study in the oil and gas industry. *Journal of Accounting and Management Information Systems, 20*(2), 264-289. https://doi.org/10.24818/jamis.2021.02004

Shahbaz, M., Karaman, A. S., Kilic, M., & Uyar, A. (2020). Board attributes, CSR engagement, and corporate performance: What is the nexus in the energy sector? *Energy Policy, 143*. https://doi.org/10.1016/j.enpol.2020.111582

Shakil, M. H. (2021). Environmental, social and governance performance and financial risk: Moderating role of ESG controversies and board gender diversity. *Resources Policy, 72*. https://doi.org/10.1016/j.resourpol.2021.102144

Silva, S. (2021). Corporate contributions to the Sustainable Development Goals: An empirical analysis informed by legitimacy theory. *Journal of Cleaner Production, 292*. https://doi.org/10.1016/j.jclepro.2021.125962

Skulstad, A. S. (2005). The use of metadiscourse in introductory sections of a new genre. *International Journal of Applied Linguistics, 15*(1), 71-86. https://doi.org/10.1111/j.1473-4192.2005.00081.x

Smeuninx, N., De Clerck, B., & Aerts, W. (2020). Measuring the Readability of Sustainability Reports: A Corpus-Based Analysis Through Standard Formulae and NLP.

International Journal of Business Communication, 57(1), 52-85. https://doi.org/10.1177/2329488416675456
Sparks, J. R., & Areni, C. S. (2007). Style Versus Substance: Multiple Roles of Language Power in Persuasion. *Journal of Applied Social Psychology, 38*(1), 37-60. https://doi.org/10.1111/j.1559-1816.2008.00295.x
Spear, S. & Roper, S. (2013). Using corporate stories to build the corporate brand: an impression management perspective. *Journal of Product & Brand Management,* 22(7), 491-501. https://doi.org/10.1108/JPBM-09-2013-0387
Stacchezzini, R., Melloni, G., & Lai, A. (2016). Sustainability management and reporting: the role of integrated reporting for communicating corporate sustainability management. *Journal of Cleaner Production, 136*, 102-110. https://doi.org/10.1016/j.jclepro.2016.01.109
Suchman, M. C. (1995). MANAGING LEGITIMACY: STRATEGIC AND INSTITUTIONAL APPROACHES. *Academy o/ Management Review, 20*(3), 571-610. https://doi.org/10.5465/amr.1995.9508080331
Talbot, D., & Boiral, O. (2018). GHG Reporting and Impression Management: An Assessment of Sustainability Reports from the Energy Sector. *Journal of Business Ethics, 147*(2), 367-383. https://doi.org/10.1007/s10551-015-2979-4
Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology, 29*(1), 24-54. https://doi.org/10.1177/0261927x09351676
Tay, D. (2021). Modelability across time as a signature of identity construction on YouTube. *Journal of Pragmatics, 182*, 1-15. https://doi.org/10.1016/j.pragma.2021.06.004
Toma, C. L., & D'Angelo, J. D. (2015). Tell-Tale Words: Linguistic Cues Used to Infer the Expertise of Online Medical Advice. *Journal of Language and Social Psychology, 34*(1), 25-45. https://doi.org/10.1177/0261927x14554484
Utz, S. (2017). Corporate scandals and the reliability of ESG assessments: evidence from an international sample. *Review of Managerial Science, 13*(2), 483-511. https://doi.org/10.1007/s11846-017-0256-x
Wood, D. J. (2010). Measuring Corporate Social Performance: A Review. *International Journal of Management Reviews, 12*(1), 50-84. https://doi.org/10.1111/j.1468-2370.2009.00274.x
Yan, B., & Aerts, W. (2014). *Rhetorical impression management in corporate narratives and institutional environment*. University of Antwerp.
Zhao, S., Liu, T., Zhao, S., Chen, Y., & Nie, J.-Y. (2016). Event causality extraction based on connectives analysis. *Neurocomputing, 173*, 1943-1950. https://doi.org/10.1016/j.neucom.2015.09.066

CHAPTER III

STUDY 2: RHETORICAL IMPRESSION MANAGEMENT IN SUSTAINABILITY REPORTING CEO LETTERS: THE EFFECT OF GENERAL AND ISSUE VISIBILITY

Abstract

Our study examines the role of visibility in the use of rhetorical IM in SR. SR has become a prevalent practice among companies, yet the narrative sections within these reports often serve as a strategic tool for companies to employ legitimization strategies. This enables managers to craft a favorable portrayal of their sustainability track record while downplaying the more problematic outcomes. Using a sample of 1,805 CEO letters in sustainability reports from international companies, this study employs multivariate regression analyses to investigate the effect of general visibility (driven by industry membership), and issue visibility (driven by sustainability controversies) on the use of rhetorical IM. The results indicate that companies tend to use rhetorical IM strategies more intensely when companies belong to industries where scrutiny and governance are lower, and when they have a higher business exposure to the environment and society. The findings of the study have important implications for understanding the communication of corporate SP and for stakeholders who rely on sustainability reports to make informed decisions.

1. Introduction

SR aims at providing information to an organization's stakeholders about its SP and commitment to social, environmental, and ethical issues. As the demand for transparency and accountability regarding SP has been increasing among stakeholders, SR has become a standard practice. According to KPMG, 80% of large and mid-cap companies around the world disclose SR (KPMG, 2020). SR is used as a communication tool providing information on companies' sustainability-related activities to address the societal call for sustainable business and to build and enhance organizational legitimacy (De Villiers & Van Staden, 2006; Hahn & Kühnen, 2013). However, many studies have demonstrated false claims, unmet promises or omitted performance outcomes in SR, which cause concerns that organizations "*talk but do not walk the sustainability*" (Boiral, 2013; Cho, Michelon, et al., 2015). As such, the credibility and the quality of sustainability-related disclosures are often debated in the business ethics literature, providing evidence that SR represents a vehicle for IM strategies (Diouf & Boiral, 2017). In the context of corporate disclosure, IM involves cherry-picking information and/or presenting it in a manner aimed at distorting readers' perceptions of firm performance (Godfrey et al., 2003; Merkl-Davies & Brennan, 2007). While the literature offers extensive empirical evidence of this practice in SR, the causal factors leading to IM in SR are not fully understood. In our study, we intend to fill this research gap by focusing on IM in SR narratives, with a specific emphasis on linguistic style as a key indicator, defining it as rhetorical IM.

Narrative sections within SR are highly discretionary, holding potential for executives to build, maintain or change corporate image and reputation, particularly through various IM strategies (Fuoli, 2018). These narratives contain personalized messages and business tales from top management, addressing key corporate events, achievements and future prospects as presented by the corporate leader (Fuoli, 2018). In particular, we build upon prior social psychology research that highlights publicity (i.e., probability of one's behavior being observed by whom and how many) as a key driver of IM (Leary & Kowalski, 1990) and explore the role of visibility notions in IM practices within the context of SR. Our investigation differentiates between general (related to industry membership) and issue (driven by sustainability controversies reflecting unethical and irresponsible business behavior) visibility, both of which hold particular relevance in sustainability research (Bowen, 2000; Dawkins & Fraas, 2010). Accordingly, we aim to answer the following research question: What is the impact of general and issue visibility on the use of IM in SR narratives?

While several authors argue that increased visibility, and hence scrutiny, makes companies less prone to IM (Delmas & Burbano, 2011; Marquis et al., 2016), others suggest that companies with higher environmental exposure and institutional pressure -and consequently, greater public visibility- are more inclined to engage in IM practices to legitimize their violations of environmental regulations and requirements (Meng et al., 2014; Stacchezzini et al., 2016). We address these contradictions in the literature in three ways. First, drawing upon the socio-political theoretical framework, we argue that general and issue visibility trigger different legitimacy-related motives, i.e., maintaining legitimacy or repairing legitimacy, which in turn influence the adoption of suitable IM strategies. In other words, the demand for legitimacy in corporate behavior leads managers to use IM strategies that match legitimacy-related requirements. The level of scrutiny for IM is clearly higher for companies operating in sensitive industries, where the institutional environment is characterized by stringent regulations and governance structures due to their higher business exposure (Aerts & Cormier, 2009; Bagnoli & Watts, 2017). For companies operating in sensitive industries, using formal language strategically and avoiding IM mechanisms are expected to help managers to improve accountability and transparency, as well as to maintain legitimacy. Issue visibility driven by sustainability controversies, on the other hand, creates a gap between the desired and current image of companies through media and news; and thus, disrupts the legitimacy status-quo

(Kuruppu et al., 2019). In such cases, managers are forced to respond to this disruption to address the spotlight of public attention directed toward companies and develop strategies to repair their damaged legitimacy (Kuruppu et al., 2019). As the relationship between general visibility and the use of IM strategies is context-dependent and is likely to be based on the company's business exposure and SP, we also assess the effect of the interaction effect of industry and controversies on the use of IM.

Second, prior studies have predominantly focused on the content-based IM, including selective disclosure (Boiral, 2013; Macellari et al., 2021; Marquis et al., 2016) or the SP-disclosure gap (García-Sánchez et al., 2022; Ruiz-Blanco et al., 2021) and little attention has been paid to the style-type, i.e., managers' use of language. In SR narratives, managers have considerable freedom to choose how the content is revealed (Du & Yu, 2020; Sandberg & Holmlund, 2015). The discretionary and voluntary nature of corporate narratives can be exploited by managers for opportunistic purposes crafting linguistic style (Stacchezzini et al., 2016; Talbot & Boiral, 2018). Style-based IM can take a variety of forms, using vague or ambiguous language, and presenting information in a favorable way. To analyze the linguistic style used in SR, we use a sample of 2,139 CEO letters from 384 companies from the EU, the UK and the US that published a stand-alone SR between 2010 and 2019. We utilize the three summary variables and the cognitive processes variables of Linguistic Inquiry and Word Count (LIWC) software, which reveal the functional, affective and cognitive linguistic elements encompassed within SR CEO letters. Through factor analysis, we identify rhetorical profiles in SR CEO letters. This approach yields three rhetorical profiles: (1) analytic; (2) assertive; and (3) defensive. Analytic relates to formal, rational and hierarchical language use (Pennebaker et al., 2015; Tay, 2021), which we refer to as no/less IM in SR, while assertive and defensive relate to narrative styles that are less technical and less factual, implying sense-giving (Bolino et al., 2008; Caliskan et al., 2021). Using these, we test our hypotheses using multivariate regression analyses. This approach allows us to take into account different linguistic style variables in our rhetorical profiles, while quantitively examining their relationship with general and issue visibility. In doing so, our analysis goes beyond the limited body of research on rhetorical IM in SR, which predominantly relies on qualitative methodologies with small sample size. In our analyses, in addition to financial, sustainability and governance control variables, we also implement an unsupervised machine learning algorithm to control for sustainability-related content intensity in the discourse, which is a novel approach in IM research.

Finally, another key factor contributing to the contradictions in the literature with respect to the drivers of IM is the lack of distinction between sections in SR. This is due to the fact that different sections of SR have distinct purposes (Fuoli, 2018). Specifically, SR narratives, such as CEO letters, tend to be more promotional in nature and include storytelling elements (Marais, 2012). By disregarding this differentiation, previous studies might have overlooked the nuanced ways in which CEOs utilize narrative sections to frame information, establish connections and respond to stakeholder's legitimacy concerns. By recognizing SR narratives' promotional nature and examining them through the lens of legitimacy concerns, our study seeks to shed light on how managers use language for IM purposes.

Our findings document that the higher general visibility the more likely companies are to use formal language. Specifically, we observe a reduced utilization of IM practices within sensitive industries. This indicates that companies in these industries tend to address stakeholder expectations regarding accountability and transparency more cautiously, aiming to establish a trustworthy reputation and maintain legitimacy given the heightened scrutiny and increased litigation risk. However, this remains true only as long as their legitimacy is not under threat. The results show that issue visibility increases due to controversies, companies in sensitive industries are more likely than their peers from non-sensitive industries to engage in defensive IM strategies. Our findings contribute to understanding the shifts in managerial concerns regarding legitimacy due to controversies, leading to changes in narrative style. Overall, our results offer valuable insights into the interconnections among visibility, scrutiny and concerns about legitimacy, enriching the scholarly discourse within the domain of business ethics.

The remainder of the paper proceeds as follows: The second part provides a brief literature review of SR and IM. In the third part, we introduce visibility notions and develop our hypotheses. The methodology used is presented in part four. The fifth part presents our research findings (including robustness checks). Finally, in part six, we discuss our findings and conclude.

2. Literature review

2.1. SR from a socio-political lens

While numerous studies have delved into the drivers behind SR practices, prior research shows that SR is considered a hybrid disclosure, in which both promotional and informative functions can be observed (Fuoli, 2018). The promotional aspects are particularly evident in narratives sections that tend to encompass more storytelling elements which are crucial in shaping the perception of stakeholders (Aerts & Cormier, 2009; Fuoli, 2018; Marais, 2012). Corporate narratives, therefore, help managers control legitimacy and promote a trustworthy company image to some extent. Managers utilize these narratives to facilitate communication between themselves and readers according to the shared conventions and expectations regarding the content and style used in recurring rhetorical situations (Fuoli, 2018). As CEO letters are, for example, social narratives between the management and stakeholders, the literature specifically examines the role of socio-political motives in shaping companies' decisions about what to and how to disclose sustainability-related information (e.g., Hahn & Lülfs, 2014).

Overall, socio-political theories (i.e., legitimacy theory, stakeholder theory, and institutional theory) view SR as a means to uphold legitimacy and address external pressures (Cho & Patten, 2007; Dawkins & Fraas, 2010; Hahn & Kühnen, 2013). According to legitimacy theory, an organization's existence is dependent on its ability to meet social expectations and its actions should therefore be perceived as desirable, proper and appropriate within social norms, values and beliefs (Suchman, 1995). Organizational legitimacy can be altered either through corporate actions or by influencing perceptions (Aerts & Cormier, 2009; Michelon et al., 2015; Ogden & Clarke, 2005). Compared to altering corporate actions, influencing perceptions may be a cheaper and easier substitute (Bansal & Kistruck, 2006). SR narratives can thus be used as a *symbolic legitimacy tool* to repair damaged legitimacy or to proactively change company image and reputation (Cho & Patten, 2007). From this perspective, SR narratives are merely a reaction to external pressures and do not always represent the organization's dedication to sustainability (Talbot & Boiral, 2018), nor is it motivated by SP or transparency and accountability concerns (Boiral, 2013; Cho et al., 2010). Within the literature diverse legitimacy strategies have been identified (i.e., pragmatic legitimacy, moral legitimacy and cognitive legitimacy) (Suchman, 2005). Marais (2012) posits that managerial focus predominantly centers on achieving moral legitimacy when communicating sustainability efforts. This approach aims to portray the organization as a good corporate citizen by employing discourse of goodwill and through invoking emotions and affection to seduce the audience. Such emphasis on moral legitimacy stems from the complex and diverse array of stakeholders, who may be interested in sustainability-related information. Similarly, stakeholder theory suggests that the needs and demands of shareholders cannot be met without satisfying the needs of stakeholders, which also implies an ever-increasing demand for socially and environmentally sustainable behavior (Michelon & Parbonetti, 2012; Prado-Lorenzo et al., 2009). Finally, institutional theory predicts that in order to attain legitimacy, organizations often respond to stakeholder pressures by implementing policies and practices that satisfy social expectations but are disconnected from internal operations (DiMaggio & Powell, 1983; Meyer & Rowan, 1977). Institutional theory literature emphasizes the role of public scrutiny and stakeholder monitoring in pushing organizations to abstain from symbolic practices in reporting (Lyon & Maxwell, 2011; Marquis et al., 2016) and align their communication with actual operations (Bromley & Powell, 2012; Delmas & Burbano, 2011).

2.2. Impression Management

The term IM stems from the social psychology literature and refers to the process by which people manage and control the image they display with the intention of changing the impression of others to be perceived and evaluated more favorably (Leary & Kowalski, 1990). In the context of organizations and corporate reporting, information may be presented in a way that affects stakeholder perceptions opportunistically to maintain or improve the corporate image (Bozzolan et al., 2015; Diouf & Boiral, 2017; Elsbach, 1994). IM in narrative disclosures can be achieved through thematic control (e.g., biased selection of themes, performance comparisons or visual presentation) and through style control (e.g., the use of complicated language to obfuscate bad news) (Merkl-Davies & Brennan, 2007). Methodologically, IM in narrative disclosures has been studied through examining the relationship between linguistic features and organizational performance (Melloni et al., 2016). Based on company performance (e.g., SP) managers can adopt assertive (i.e., proactively trying to change company image, build legitimacy, reliability and reputational characteristics) and/or defensive (i.e., responding to a threat or damage to company image) strategies (Aerts & Yan, 2017; Boiral et al., 2020; Yan et al., 2019).

When company activities and outcomes are desirable, managers tend to adopt IM as an assertive strategy in a proactive manner to enhance and emphasize companies' positive outcomes to gain competitive advantage (Aerts & Cormier, 2009; Barkemeyer et al., 2014;

Boiral et al., 2020). Assertive IM involves highlighting unique organizational competencies and business skills (Boiral et al., 2020; Bolino et al., 2008). Managers using this strategy stress the importance, relevance and scope of positive environmental outcomes or actions positing an acclaiming and self-confident stance in the message delivered (Aerts & Cormier, 2009; Aerts & Yan, 2017). In other words, the aim of assertive IM is to make a positive outcome more obvious to the audience and project an impression of organizational success. For example, Talbot and Boiral (2018) identify optimistic neutralization techniques (i.e., self-proclaimed excellence and promotion of systemic view), which exaggerate the actual commitment to sustainability and climate change issues. These strategies are reflected in managers' use of language, where a stronger positive tone, extensive self-referencing, emphatic certainty expressions and achievement- and future-related content references are used (Aerts & Yan, 2017).

On the other hand, in circumstances that threaten legitimacy, managers employ defensive IM strategies (Bolino et al., 2008). In the literature, defensive IM strategies, which encompass distractions, apologies, excuses, justifications or self-handicapping, are used as an umbrella term for reactive mechanisms (Boiral et al., 2020; Kibler et al., 2021)[10]. By using such techniques, managers can deny any responsibility for negative actions or outcomes, or seek to re-establish a positive identity by removing negative perceptions and giving sense to the organization's actions (Caliskan et al., 2021; Tata & Prasad, 2015). They may also prefer to withhold or obfuscate unfavorable performance-related information, as it may cause damage (Dye, 1985; Fabrizio & Kim, 2019; Verrecchia, 1983). As such, within the academic literature,

[10] Drawing upon Tedeschi and Melburg's (1984) IM framework, we classify IM as assertive versus defensive. In our study, we use the framework of Tedeschi and Melburg (1984), as it is well suited for legitimacy theory (Ogden & Clarke, 2005) and it facilitates a comprehensive understanding for IM with its diverse subcategories. There are, however, different literature streams that define and classify IM in a different manner. Hooghiemstra (2000) named these IM strategies as acclaiming and accounting; Bansal and Kistruck (2006) classified IM strategies as demonstrative versus illustrative; Elsbach (1994) classified IM strategies as accommodative versus defensive; and Higgins and Walker (2012) classified IM strategies as logos, ethos and pathos.

a prevailing view is that a poor SP tends to prompt organizations to resort to IM strategies as a means of safeguarding their legitimacy. For example, Boiral (2016) and Talbot and Boiral (2018) qualitatively investigate IM in SR of mining and energy firms, which are used to neutralize and legitimize the possible impacts of companies' actions on biodiversity and to justify or conceal poor SP. Similarly, Hahn and Lülfs (2014) identify six legitimizing strategies for reporting negative SP in SR, by conducting a qualitative analysis of 40 SR of companies. For defensive IM, managers use more cautious framing and sense-giving devices, such as engagement markers (e.g., "consider", "note that", "you can see that") that "*project an aura of credibility gained by openness*" and hedging expressions (e.g., "might", "perhaps", "possible") which help creating distance from the message and avoid direct responsibility (Aerts & Yan, 2017, p. 416; Hyland, 1998). Defensive strategies have been also associated with the remedial use of causal language, which helps to portray negative results as understandable and minimize management responsibility for them (Zhang & Aerts, 2015). We highlight that both assertive or defensive IM strategies are not supposed to be perceived as negative strategies, as they both serve self-presentation and information-sharing purposes (Yan et al., 2019). Despite the extensive research on IM in SR, prior studies based on small samples are inadequate to explore the causal factors for using rhetorical IM strategies in SR in a broad set of industries and companies, which have hindered a comprehensive grasp of the phenomenon.

3. Hypotheses development

Social psychology research identifies publicity as one of the main drivers of IM (Leary & Kowalski, 1990). Higher publicity increases the perceived risk and leads to more concerns about how others see one's behavior (Leary & Kowalski, 1990). This resonates with the notion of visibility which is the degree to which something is seen or known by the public. With regard to organizations, visibility escalates institutional pressure and public scrutiny over corporate behavior. As such, it is believed to determine companies' tendency to symbolism and IM (Greenwood et al., 2011). The literature shows that the social and environmental visibility of an organization is driven by two sets of factors: (1) general organizational characteristics (i.e., general visibility); and (2) organizations' proximity to a given issue resulting from actual SP (i.e., issue visibility) (Bowen, 2000; Dawkins & Fraas, 2010). General visibility has been mainly linked to organizations' size and industry membership (Alonso-Almeida et al., 2015; Ruiz-Blanco et al., 2021), and is associated with stringent scrutiny over companies' strategies and disclosures, yet it does not create a legitimacy threat. Disruption in the "legitimacy status-quo" can be expected as a result of a major sustainability controversy that increases the issue visibility of companies.

3.1. General visibility driven by sensitive industry membership and IM

Operating in a sensitive industry is one of the initial impressions one may have of a company (Aerts & Cormier, 2009). Companies operating in these "dirty" industries, such as oil and gas extraction companies, have greater business exposure, that is the degree to which an organization affects its environment (Michelon et al., 2013)[11]. These industries face substantial scrutiny from a broad range of stakeholders and are constantly under the spotlight in case of a controversial event (Patten, 2002). Prior research shows that to project a more transparent and credible image in response to greater scrutiny by a wide range of stakeholders, companies operating in sensitive industries tend to disclose higher-quality sustainability-related information (Cho & Patten, 2007; Nilipour et al., 2020; Villiers & van Staden, 2011).

As institutional theory argues, accountability and reporting practices become ritualized in a more scrutinized environment, mainly due to coercive isomorphism found in shared agreements such as common industry regulations (Larrinaga-González, 2007; Weick, 1995).

[11] Prior literature consider materials, oil & gas and utilities as sensitive industries (Cho et al., 2012; García-Meca & Martínez-Ferrero, 2021).

Managers tend to imitate the decisions of industry leaders that are in the same institutional reference groups, to appeal legitimate through using institutional templates (DiMaggio & Powell, 1983; Van Caneghem & Aerts, 2011). Such conditions lead to increasingly standardized and rationalized practices in organizations, where similar replicable and easily defensible narratives can be seen, which result in more rigid explanations (Aerts & Tarca, 2010). Furthermore, the presence of an evaluative audience may lead to a low-risk disclosure attitude in corporate narratives, as well as to follow more injunctive rather than descriptive norms in disclosures. In line with this argumentation, Aerts et al. (2013) document that in highly scrutinized institutional environments, managers tend to use technical accounting explanations that are based on formal and analytical language and avoid causal expressions, which are discretionary and more open to IM. Thus, we argue that industry-level institutional settings may enhance the preference for formal and rigid language use with specialized terminology and inherent calculative relationships.

Higher scrutiny, more stringent regulatory conditions and higher litigation risk in sensitive industries are found to lead to a more boilerplate language with extensive elaboration on firm performance and to discourage managers from using IM strategies (Aerts & Cormier, 2009; Bagnoli & Watts, 2017; Lyon & Maxwell, 2011). As scrutiny increases the likelihood and expected costs of being caught in the attempt to obfuscate or cover up negative outcomes, IM strategies can backfire and be perceived as misleading and/or cheap talk (Kim et al., 2004; Rogers et al., 2011). Similarly, Marquis et al. (2016) provide empirical evidence that companies causing higher environmental damage are less likely to selective disclosure if they are subject to increased monitoring by civil society organizations. In such case, being transparent and proactive help maintain legitimacy (Kuruppu et al., 2019). Thus, it is expected that managers of companies that operate in sensitive industries will curb their tendency to use both assertive and defensive IM strategies to avoid litigation risks.

In contrast, in non-sensitive industries companies may stay under the radar. Crilly et al. (2012, p. 1436) argue that companies in non-sensitive industries *offer high potential for information asymmetry and can easily build smokescreens around their internal practices*. Less stringent regulations and less public scrutiny leave companies operating in non-sensitive industries in a more discretionary environment. We, therefore, predict that general visibility will deter companies from sensitive industries from pursuing (both assertive and defensive) IM strategies and enhance the use of formal language[12]:

H1a: General visibility driven by sensitive industry membership is positively associated with the use of more formal language.

H1b: General visibility driven by sensitive industry membership is negatively associated with the use of rhetorical IM.

3.2. Issue visibility driven by sustainability controversies and IM

The issue visibility of an organization is driven by the company's ESG performance. Corporate controversies, defined as *publicly observable events that are expected to have negative implications on the firm* (Del Giudice & Rigamonti, 2020, p.2), attract media attention and create a gap in the organizational image, jeopardizing the firm's legitimacy and value (DasGupta, 2021; Elsbach & Sutton, 1992). As the discrepancy between the desired and current image of companies increases, managers tend to get more motivated to use IM strategies (Leary & Kowalski, 1990). Due to the escalating prevalence of adverse corporate incidents in the form of corporate controversies, legitimacy techniques may be utilized to repair and restore the damaged legitimacy (O'Donovan, 2002). For example, prior research identified

[12] While previous literature frequently cites size as a proxy for general organizational visibility, this approach has an inherent weakness (Bowen, 2000). Larger companies may not only be more visible but may also have more resources to allocate towards sustainability and reporting. Accordingly, authors propose an alternative perspective, suggesting that industry membership could serve as a viable measure to assess general visibility (Patten, 1991; Reverte, 2009; Ruiz-Blanco et al., 2021).

communicative legitimization and neutralization strategies that organizations follow when disclosing negative sustainability outcomes (e.g., Boiral, 2016; Hahn & Lülfs, 2014). These strategies are linked to IM, which helps managers disassociate themselves from negative sustainability-related events through denying responsibility for negative outcomes or rationalizing and legitimizing ethically questionable behaviors (Hahn & Lülfs, 2013).

Issue visibility has been associated with sustainability-related incidents or controversies that attract high levels of negative media attention (Hooghiemstra, 2000). Poor performance outcomes, such as corporate controversies that are reflected in media and news (e.g., the BP oil spill in the Gulf of Mexico) can create a challenging environment to use an assertive style and trigger defensive IM behavior. Hence, such controversies may restrict managers from proactively disclosing positive activities and outcomes, as the companies' SP is not desirable and threatens legitimacy due to more political and social pressures. Instead, controversies would lead managers to adopt defensive IM strategies to justify actions or obfuscate poor SP-related outcomes. On the other hand, in case of no controversy, managers have a greater space to proactively emphasize the positive activities and outcomes to gain competitive advantage and to create a desired image. Thus, we propose the following hypotheses:

H2a: Issue visibility driven by sustainability controversies is positively associated with defensive IM.

H2b: Issue visibility driven by sustainability controversies is negatively associated with assertive IM.

Once controversial events occur for companies (across both sensitive and non-sensitive industries), corporate scandals spread quickly through the media, potentially leading to long-term reputational and financial harm (Aouadi & Marsat, 2018; Del Giudice & Rigamonti, 2020)[13]. In the presence of a controversy, companies exhibit a decreased level of cautiousness for accountancy and transparency and adopt strategies to repair their legitimacy as a response to the threat and heightened visibility given that stakeholders demand efforts and communication about resolving the occurred controversies and disrupted legitimacy status-quo (Kuruppu et al., 2019). As such, despite the diminishing effects of higher scrutiny and stringent regulatory conditions of sensitive industries on the use of IM, controversies can force companies across all industries to shift their strategic narratives and messaging mainly due to increased legitimacy concerns. This is also in line with prior research which demonstrates the use of legitimization strategies for negative incidents among sensitive industries (e.g., chemical, oil and gas) (Hahn & Lülfs, 2014; Talbot & Boiral; 2015). Consequently, the relationship between general visibility and IM is expected to be altered because of the effect of issue visibility led by controversies. Moreover, given the challenges managers face in highlighting any favorable results related to their sustainability initiatives during times of controversies, our projection is that issue visibility will amplify the resistance of sensitive industry members to use an assertive style. As reported by Rogers et al. (2011), overly optimistic language (as in assertiveness) can increase litigation risk under such circumstances. Thus, we expect that:

H3a: The negative relationship between general visibility and defensive IM is positively moderated by issue visibility.

H3b: The negative relationship between general visibility and assertive IM is negatively moderated by issue visibility.

[13] Both Aouadi and Marsat (2018) and Del Giudice and Rigamonti (2020) show that controversies can occur all across industries, i.e., in both sensitive and non-sensitive.

4. Methodology

To study rhetorical IM in SR, we employ a three-step methodological approach. First, we conduct an automated text analysis to examine the linguistic style used in CEO letters in SR. CEO letters are used as an important communication channel between managers and stakeholders (Amernic & Craig, 2007). Through CEO letters, managers aim to convince stakeholders regarding a company's legitimacy and improve confidence in the organization (Fuoli, 2018; Jonäll & Rimmel, 2010). CEO letters have been widely used in prior IM research, as they are highly discretionary vehicles and subject to IM (e.g., Barkemeyer et al., 2014; Bozzolan et al., 2015; Im et al., 2021; Caliskan et al., 2021). Second, we conduct a factor analysis to reveal co-occurrence patterns among linguistic markers derived from the previous step to identify IM strategies used in SR. Finally, we test our hypotheses with the proposed empirical models.

4.1. Sample selection and data collection

The sample selection process was conducted in two steps. We select listed companies from the EU, the UK and the US that published a stand-alone SR for the fiscal year 2016 according to the Refinitiv (Thomson Reuters'Asset4ESG) database. We extend the time frame for the same companies from 2010 to 2019 and manually collect the reports in PDF from their websites, corporate register (corporateregister.com), and GRI global reporting (globalreporting.org) databases. We do not include companies that operate in the healthcare, financials, real estate, government activities, and academic and educational services sectors; companies that do not publish their reports in English; companies that use integrated reporting; SR that do not include CEO letters; SR that are protected and cannot be processed; and CEO letters counting fewer than 350 words[14] in length. We manually processed CEO letters in SR

[14]The Receptiviti API generates scores on language using different measures, such as personality, emotions or cognition. In creating measures, the Receptiviti API uses samples that exceed 350 words for baselining its

in accordance with prior research (Clarkson et al., 2008; Cong et al., 2014; Fehre & Weber, 2016; Na et al., 2020), and excluded CEO letters that do not contain sustainability-related information. This selection process yielded a sample of 2,139 CEO letters from 384 companies that we later use in factor analysis to create rhetoric variables. Table 1 shows our selection criteria.

Table 1: Selection criteria

Step 1: Criteria of selection of companies	Excluded amount	Remaining number of companies
SR 2016	-	2,649
Country of headquarters – Europe, US, UK	(1,386)	1,263
Excluding non-EU	(85)	1,178
Excluding healthcare, financials and real-estate sectors	(293)	885
Excluding companies with missing financial or sustainability data on the database	(306)	579
Manual download of SR	(171)	408
Step 2: Criteria for CEO letter selection	**Excluded amount**	**Remaining number of CEO letters for 10 years**
Expected number of SR CEO letters in 10 years	-	4,080
Excluding integrated reporting formats	(474)	3,607
Excluding SR with no CEO letters	(390)	3,217
Excluding SR with no CEO letters in English	(16)	3,201
Excluding CEO letters with less than 350 words	(704)	2,496
Excluding SR CEO letter not able to process (e.g., secured pdf files)	(76)	2,420
Excluding SR CEO letter with no sustainability information	(281)	2,139
		(from 384 companies)
Step 3: Matching Refinitiv and Bloomberg data		
	-	2,139
Excluding companies that do not match with Bloomberg data	(334)	1,805

measures (https://docs.receptiviti.com/). Based on this, we impose minimum word limits for our investigation, that is documents that have less than 350 words are too short to be interpreted for our analysis.

Matching CEO letters
1,805 (from 345 companies)

We combined the sample of CEO letters with data we needed for explanatory and control variables derived from Refinitiv and Bloomberg and excluded observations with missing data. The final sample used in the multivariate analyses consists of 1,805 CEO letters from 345 companies headquartered in 18 countries. Table 2 presents the final sample summary statistics with respect to distribution by year, country, region, and industry.

Table 2: Sample summary statistics

Panel A Sample by country and year											
Country	Year										
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total
Austria	1	1	2	3	2	2	2	1	2	2	18
Belgium	2	2	2	1	3	2	3	2	2	2	21
Denmark	2	3	3	4	5	2	5	4	4	5	37
Finland	5	4	6	6	5	5	3	3	3	3	43
France	9	9	9	5	7	7	5	5	5	6	67
Germany	2	3	3	2	2	3	4	12	16	13	60
Greece	3	0	3	3	2	5	3	4	4	3	30
Hungary	1	0	0	0	0	0	0	1	1	1	4
Ireland	0	2	2	2	2	1	2	2	3	2	18
Italy	4	5	6	5	7	6	6	5	5	6	55
Luxembourg	1	2	3	3	3	2	1	2	2	1	20
Netherlands	5	7	7	6	5	5	3	2	2	2	44
Poland	0	0	1	1	0	1	2	2	1	3	11
Portugal	1	1	2	1	1	1	2	1	2	1	13
Spain	3	4	5	5	4	4	4	5	4	4	42
Sweden	8	10	11	13	12	13	12	12	12	10	113
UK	22	26	22	32	37	34	35	28	35	34	305
US	52	69	70	88	83	91	107	107	114	123	904
Total	121	148	157	180	180	184	199	198	217	221	1,805
Panel B Sample by region and year											
Region	Year										
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total
EU	47	53	65	60	60	59	57	63	68	64	596
UK	22	26	22	32	37	34	35	28	35	34	305
US	52	69	70	88	83	91	107	107	114	123	904
Total	121	148	157	180	180	184	199	198	217	221	1,805
Panel C Sample by industry and year											
Industry [a]	Year										
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total
Cons. Disc.	6	9	7	6	8	9	9	12	17	19	81
Cons. Staple	16	22	22	27	23	24	31	23	26	24	238
Indust. Prod.	10	13	15	20	22	20	24	26	30	30	210
Indust. Ser.	15	21	16	20	25	23	23	22	20	19	204
Materials	16	20	29	32	28	29	31	34	37	39	295
Media	7	6	7	6	8	9	9	7	8	9	76
Oil & Gas	13	15	19	19	21	21	20	23	20	21	192
Ren. Energy	0	0	0	0	0	0	0	0	1	1	2
R&Wstaples	8	9	8	9	8	6	9	10	12	9	88
R&Wdisc.	5	6	6	10	8	11	10	8	13	10	87
Software	4	3	3	4	3	3	5	5	5	5	40
Hardware	2	1	2	1	1	3	2	3	1	3	19
Telecom.	7	7	5	8	4	6	4	7	5	7	60
Utilities	12	16	18	18	21	20	22	18	22	25	192
Total	121	148	157	180	180	184	199	198	217	221	1,805

Notes. [a] Based on Bloomberg Industry Classification Standard (BICS) level 2. *Abbrevation.* Cons. Discr., Consumer Discretionary Products; Cons. Stapl., Consumer Staples Products; Indust. Prod., Industrial Products; Indust. Ser., Industrial Services; Ren. Energy, Renewable Energy; R&Wstaples, Retail & Wholesale – Staples; R&Wdisc., Retail & Wholesale – Discretionary; Software, Software & Tech Services; Hardware, Tech Hardware & Semiconductors; Telecom., Telecommunications.

4.2. Linguistic analysis

To examine the linguistic style used in CEO letters, we use LIWC, which is based on strong empirical evidence that language can provide rich insights into people's psychological states, including emotions, thinking styles, and social concerns (Boyd et al., 2022). This software has been extensively tested in numerous social psychology studies (Tausczik & Pennebaker, 2010) and has also been used in corporate reporting context (e.g., Aerts & Yan, 2017; Merkl-Davies & Brennan, 2017). We use LIWC-22 to reveal the functional, affective and cognitive linguistic components of CEO letters in SR. We employ the three summary psychosocial variables of LIWC*:* (1) *analytical thinking;* (2) *clout*; and (3) *emotional tone.* In addition to these, we use the *cognitive processes* variable.

The three summary variables are composite measures derived from previously published findings from Pennebaker Language Lab and converted to percentiles based on standardized scores from large comparison samples (Boyd et al., 2022). Each of the variables is transformed into a scale from 1 to 100, and they are mainly based on function words. Function words have been found to be reliable markers of psychological states, revealing how people are thinking (Pennebaker et al., 2014). For example, high use of single-person or second/third-person pronouns have been associated with self- or other-focus, respectively; auxiliary verbs with the use of passive language; articles with formal and structured style of writing; and conjunctions with cognitive complexity (Aerts & Yan, 2017; Hyland, 2005; Pennebaker et al., 2007). The *analytical thinking* variable captures *"the degree to which people use words that suggest formal, logical and hierarchical thinking patterns"* (Pennebaker et al., 2015; Tay, 2021). The lexical categories of higher *analytical thinking* score include articles, prepositions which reflect a more technical language; while a lower score includes pronouns, auxiliary verbs, conjunctions and adverbs which reflect a more informal and personal language (Pennebaker et al., 2015; Tay, 2021). *Clout* variable refers to the relative social status, confidence, or leadership that people display through their language (Boyd et al., 2022). A higher *clout* score is measured by 1st person plural (*we*) and second-person pronouns (*you*), which reflects confident, credible and collectively-oriented language; while a lower clout score is measured by tentative words (e.g., *maybe*, *perhaps*) reflecting tentativeness, humble and anxious language (Kacewicz et al., 2014; Pennebaker et al., 2015; Tay, 2021). Finally, *emotional tone* variable puts positive tone and negative tone dimensions into a single variable (Cohn et al., 2014). The higher the score, the more positive the tone is. Tone as a measure of text sentiment has been widely used as a proxy of disclosure balance in SR (Cho et al., 2010; Muslu et al., 2019). Table 3 summarizes the indicators of analytical thinking, clout, and emotional tone, their directions and example words.

Table 3: LIWC summary variables and their indicators

Summary variable	Direction of effect	Indicator	Examples of words
Analytic thinking	+	Article, preposition	A, an, the, of, in, for
	-	Pronouns, auxiliary verbs, adverbs, conjunctions, negations	I, it, be, have, just, about, but
Clout	+	First plural and second person pronouns, positive tone words	We, you, they, good, well, new
	-	First singular pronouns, tentative	I, if, or, maybe, perhaps
Emotional tone	+	Positive tone words	Good, well, new, happy, love
	-	Negative tone words	Bad, wrong, much, hate

Sources : (Boyd et al., 2022 ; Cohn et al., 2004 ; Kacewicz et al., 2014 ; Pennebaker et al., 2014)

Cognitive processes is a composite variable that includes markers of cognitive complexity such as insight words, causation words, discrepancy words, tentative words, certitude words and differentiation words. Cognitive processes words signal thought, causality and insight, and are used in situations, *when people wish to transmit facts, reconstruct events and provide explanations for them* (Brownlow et al., 2020, pp. 11-12). As such, they were found to be used in accounting narratives for sense-giving purposes (Aerts & Yan, 2017; Merkl-Davies & Brennan, 2007), which relate to both justifications and explanations, as one way of defensive IM strategies (Boiral et al., 2020). The variable is expressed as a percentage of the total words used in any given language sample. Table 4 presents components of the cognitive processes variable with their most frequently used exemplars.

Table 4: Components of the cognitive processes variable

Cognitive processes		
Component variable	Function	Example words
Insight	Engagement marker	Know, how, think, feel
Causation	Causal reasoning	How, because, make, why
Discrepancy	Directive language	Would, can, want, could
Tentative	Linguistic hedging	If, or, maybe, perhaps
Certitude	Boosters	Really, actually, of course, real
Differentiation	Exclusion words	But, not, if, or

Source: (Boyd et al., 2022)

4.3. Factor analysis

We follow prior literature (Aerts & Yan, 2017; Pennebaker et al., 2014) and use LIWC-derived variables as inputs to factor analysis to identify salient linguistic structures in written texts. The identified constructs serve further as our test variables. Hence, we employ principal component factor analysis with varimax orthogonal rotation. Table 5 shows the variables and their respective factor loadings.

Table 5: Factor analysis on linguistic markers (N=2,139)

Variable	Factor 1	Factor 2
Factor loading (orthogonal varimax)		
Analytic thinking	-0.438	-0.718
Emotional tone	0.833	-0.053
Clout	0.846	0.142
Cognitive processes	-0.098	0.893

The analysis led to the identification of two uncorrelated factors with eigenvalues greater than 1.0, which cumulatively explain 72% of the overall variance. They group the stylistic and content characteristics that tend to co-occur in CEO letters in SR, revealing the main rhetorical profiles in our sample. We use a cut-off of 0.20 for interpretation purposes and label the factors as follows: assertive (Factor 1) and defensive (Factor 2).

Both factors have negative loadings on analytic thinking. This provides strong support that analytic indicates a more formal and analytical style with minimal audience engagement, while identified factors represent highly narrative, less technical and less factual writing styles. Factor 1 is based on positive emotional tone and clout, highlighting social connections and accomplishments with the use of optimistic narratives, as we refer to as *assertive*. Factor 2 is balanced communication (e.g., neutral loading of emotional tone) with high use of cognitive processes markers that cautiously provide explanation and meaning to organizational outcomes, as we refer to as *defensive*. Within Factor 2, the loadings suggest that this style provides complex meanings, justifications and explanations for organizational events while avoiding strong emotional bias.

To ensure accurate factor labeling and strengthening the robustness of our conclusions, we additionally accounted for the correlations between the factor variables and linguistic style and content variables on LIWC-22 (Table 6). Analytic is highly correlated with articles ($r = .517$) and prepositions ($r = .423$), which connotate conceptual signaling and relationships within language; conversely a lower prevalence of articles and prepositions tends to indicate a more narrative language style (Jordan et al., 2019). Additionally, analytic features a negative correlation with personal pronouns ($r = -.657$), suggesting a formal language style with minimal audience engagement. Numbers, positively correlated with analytic ($r = .219$) and negatively correlated with assertive and defensive ($r = -.211$; $r = -.275$, respectively), indicate a factual and hard disclosure style. Furthermore, allure, a group of words that is used in advertisements and persuasive communications, auxiliary verbs, conjunctions and adverbs demonstrate a negative correlation with analytic ($r = -.443$, $r = -.534$; $r = -.373$; $r = -.496$, respectively), which indicates a less narrative and storytelling language (Boyd et al., 2022; Tausczik & Pennebaker, 2010; Pennebaker et al., 2014).

Conversely, assertive features a negative correlation with articles and prepositions (r = -.505; r = -.052) and positive correlation with personal pronouns (r = .695), indicating a more involved and personalized and narrative communication approach. Assertive tendencies also associate significantly with affiliation (r = .754), achievement (r = .458) and positive tone (r = .777), reflecting a language style that highlights social connections and accomplishments. Moreover, assertive displays positive correlations with sustainability-related words, i.e., human rights, employee, social community, environment (Pencle & Malaescu, 2016) (r = .296; r = .325; r = .309; r = .109, respectively) and allure (r = .351), suggesting a tendency to present optimistic narratives as in advertisements in discussing these aspects.

On the other hand, defensive exhibits negative correlations with linguistic markers like articles (r = -.176) and prepositions (r = -.142), and displays positive correlations with personal pronouns (r = .309), suggesting a more narrative style similar to assertive. Unlike assertive, however, defensive style is neutral in tone, showing no significant correlations with positive or negative tone or emotions. It also shows positive correlations with auxiliary verbs, conjunctions, adverbs and allure (r = .601; r = .304; r = .554; r = .389, respectively). Among these variables only conjunctions are significantly correlated with assertive (r = .286). This indicates that storytelling and a narrative language style is stronger in defensive style (Tay, 2021). Furthermore, defensive correlates significantly with negation words, verbs, cognitive processes words and *need* and *want* states (r = .472; r = .670; r = .886; r = .252; r = .251, respectively), which indicates that this style includes descriptions related to a past events, self-evaluation, i.e., ex-post disclosure, as well as actions that are required for a better performance, i.e., ex-ante disclosure (Akstinaite et al., 2020; Boyd et al., 2022; Tausczik & Pennebaker, 2010).

In acknowledging the diverse interpretations and classifications within the field of IM, it is imperative to note that the terms “assertive” and “defensive” encompass a broad spectrum of subcategories in the literature, including enhancements, entitlements, excuses and justifications (Tedeschi & Melburg, 1984). In our investigation, as indicated by the correlations observed, the assertive category is notably associated with what might be considered as enhancements, portraying an overly optimistic perspective, while the defensive category aligns more with justifications or the act of sensemaking and rationalizing the communicated content. It is crucial to emphasize that our characterization of analytic, assertive and defensive factors is specific and tailored to this context of SR, potentially holding nuanced differences from IM descriptions within traditional FR narratives. The complexity of sustainability information, being multidimensional and predominantly qualitative in nature, as opposed to financial information, allows the creation of diverse sustainability narratives.

Table 6: Correlation matrix between identified factors and linguistic indicators

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
(1) Analytic	1.000																							
(2) Assertive	-0.438*	1.000																						
(3) Defensive	-0.718*	0.000	1.000																					
(4) Articles	0.517*	-0.504*	-0.177*	1.000																				
(5) Prepositions	0.423*	-0.052	-0.142*	0.354*	1.000																			
(6) Personal pronouns	-0.657*	0.695*	0.309*	-0.539*	-0.190*	1.000																		
(7) Numbers	0.219*	-0.211*	-0.275*	-0.153*	-0.200*	-0.255*	1.000																	
(8) Allure	-0.443*	0.351*	0.390*	-0.142*	0.036	0.368*	-0.189*	1.000																
(9) Auxiliary verbs	-0.534*	0.084*	0.609*	0.040	0.034	0.141*	-0.273*	0.464*	1.000															
(10) Conjunctions	-0.374*	0.287*	0.305*	-0.183*	0.071*	0.140*	-0.325*	0.145*	0.084*	1.000														
(11) Adverbs	-0.496*	0.069*	0.554*	0.016	0.055	0.115*	-0.201*	0.362*	0.549*	0.198*	1.000													
(12) Affiliation	-0.535*	0.754*	0.217*	-0.516*	-0.046	0.883*	-0.215*	0.344*	0.063*	0.250*	0.051	1.000												
(13) Achievement	-0.073*	0.458*	-0.073*	-0.160*	0.078*	0.228*	-0.157*	0.132*	-0.044	0.140*	-0.093*	0.334*	1.000											
(14) Positive tone	-0.238*	0.777*	-0.048	-0.279*	0.038	0.369*	-0.199*	0.298*	0.051	0.239*	0.035	0.470*	0.516*	1.000										
(15) Negative tone	0.030	-0.187*	0.029	0.029	0.020	-0.065*	-0.020	-0.055	0.013	0.039	0.039	-0.059*	-0.080*	-0.037	1.000									
(16) Human rights	-0.078*	0.293*	-0.069*	-0.211*	-0.046	0.201*	-0.069*	0.005	-0.217*	0.228*	-0.198*	0.335*	0.183*	0.287*	0.072*	1.000								
(17) Employee	-0.078*	0.325*	-0.055	-0.210*	-0.067*	0.207*	-0.109*	0.011	-0.219*	0.258*	-0.226*	0.323*	0.281*	0.302*	0.023	0.794*	1.000							
(18) Social community	-0.078*	0.309*	-0.019	-0.207*	-0.059*	0.175*	-0.155*	0.042	-0.181*	0.245*	-0.196*	0.330*	0.337*	0.297*	-0.015	0.620*	0.659*	1.000						
(19) Environment	0.063*	0.109*	-0.052	-0.179*	-0.032	-0.017	-0.076*	-0.175*	-0.207*	0.155*	-0.215*	0.080*	0.143*	0.096*	-0.016	0.254*	0.357*	0.526*	1.000					
(20) Negate	-0.331*	-0.088*	0.472*	0.033	-0.041	0.022	-0.117*	0.219*	0.418*	0.108*	0.427*	-0.043	-0.114*	-0.065*	0.111*	-0.140*	-0.169*	-0.114*	-0.108*	1.000				
(21) Verb	-0.591*	0.224*	0.670*	-0.069*	0.072*	0.287*	-0.319*	0.610*	0.867*	0.162*	0.549*	0.232*	0.025	0.183*	0.007	-0.139*	-0.146*	-0.082*	-0.186*	0.397*	1.000			
(22) Cognitive processes	-0.369*	-0.099*	0.894*	0.040	0.113*	0.076*	-0.286*	0.313*	0.499*	0.226*	0.441*	0.054	-0.031	-0.015	-0.007	-0.092*	-0.062	-0.012	0.000	0.414*	0.576*	1.000		
(23) Need	-0.127*	-0.038*	0.253*	0.028	0.019	-0.025	-0.158*	0.131*	0.243*	0.134*	0.174*	-0.020	-0.028	0.009	0.183*	-0.032	-0.048	0.045	0.018	0.206*	0.296*	0.259*	1.000	
(24) Want	-0.227*	0.150*	0.251*	-0.046	-0.003	0.191*	-0.137*	0.181*	0.171*	0.086*	0.209*	0.134*	0.039	0.127*	-0.054	-0.017	0.032	-0.018	-0.037	0.153	0.251*	0.229*	0.055	1.000

*Correlations significant at p<0.01

4.4. Empirical models

To test our hypotheses, we employ the following regression model (full model):

Rhetorical style = *f*(INDsen, ESGcontro, INDsen X ESGcontr, controls)

Our dependent variable represents three different rhetorical styles. First, we use analytical thinking variable, that is negatively loaded in the factor analysis for assertive and defensive factors, which indicates formal language use or less/no IM. The other two dependent variables are derived from our factor analysis: (1) assertive (Factor 1); and (2) defensive (Factor 2). These two variables correspond to writing styles that are predominantly narrative and include storytelling and IM elements.

The first variable of interest (INDsen), being a dummy variable that indicates a company's membership in sensitive industries, captures general visibility (H1 and H3). In defining the variable, we follow prior literature (Cho et al., 2012; García-Meca & Martínez-Ferrero, 2021) and consider Materials, Oil & Gas and Utilities as sensitive industries (dummy variable is coded one for those industries)[15]. The second variable of interest (ESGcontro) is the ESG controversies score as a proxy for issue visibility (H2 and H3). The score is calculated by Refinitiv based on 23 ESG controversy topics and is captured from global media (third-party sourced information) that materially impact the corporations. The score allows comparison across industries and resolves the market cap bias, which large companies suffer from, as they attract more media attention than smaller companies through adjusting and normalizing scores based on company size and industry (Shakil, 2021; Vasilescu & Wisniewski, 2020). The controversies score is expressed as a percentage rank, where a higher score implies fewer controversies. For a more straightforward interpretation, we reverse code the controversies

[15] We acknowledge that all companies in all industrial sectors can conduct activities with potential implications for the environment and society. Our selection of "dirty industries", however, include the industries with the highest business exposure.

score (i.e., companies with controversies is equal to 1; and no controversies is equal to 0). To test the industry moderation effect (H3) we include the interaction term between INDsen and ESGcontro.

Our models also include a number of control variables. First, we acknowledge the distinction between corporate responsibility, positive social and environmental performance and irresponsibility reflected by controversies (Riera & Iborra, 2017; Strike et al., 2006). As they can co-occur (companies can be responsible and irresponsible at the same time), potentially having different effects on corporate disclosure behavior (Dawkins & Fraas, 2010; Meng et al., 2014), including rhetoric found in SR CEO letters, we control for environmental and social performance. In doing so, we use the mean of Refinitiv's environmental and social scores (EnvSoc), following prior literature (e.g., Ioannou & Serafeim, 2012; Michelon et al., 2015). Furthermore, we acknowledge that language is not an independent stylistic device but is also related to the discourse in the text (Hyland, 1998) and IM practices may change in different sustainability pillars as indicated in prior literature for selective disclosure (Roszkowska-Menkes et al., 2024). Hence, we control for the content of the CEO letter. We account for that by identifying discourses present in the examined documents through employing topic modeling based on the latent Dirichlet allocation (LDA) (a detailed explanation of the procedure can be found in Appendix 1). This approach allows us to control for sustainability-related content intensity and business strategic-case-related content intensity. As the discourses represent probability and their sum equals one, there is a perfect inverse correlation (-1.0) between them. Thus, we use only one of the variables (strategic business case-related content intensity), namely StratDis, as a control in our regression models. To control for organizational characteristics and corporate governance mechanisms that may influence IM behavior, we use GRI dummy variable indicating whether the company follows the GRI standards/guidelines in its disclosure process (Chelli et al., 2018); assurance dummy variable (ASSU) indicating whether company's report was subject to independent assurance (Braam et al., 2016); corporate governance variables including CSR committee (CSRcom) (Amran et al., 2014), percentage of female directors on board (BRDwmn) (Ben-Amar et al., 2017), board size (BRDsize), percentage of independent board directors (BRDind) (Frias-Aceituno et al., 2014) and CEO duality (CEOdual) (Helfaya & Moussa, 2017). As firm-level financial controls, we use ROA as a proxy of a company's profitability, and the natural logarithm of total assets as an indicator of the company's size (Size). Additionally, we add dummy variables to control for the region (UK and US), where the company is headquartered, since previous studies have found that rhetorical IM is influenced by the country-level institutional environment (Aerts & Yan, 2017; Jackson & Apostolakou, 2010). This allows us to control for the regional-level institutional settings, that may play a significant role shaping IM practices. As Matten and Moon (2008) argue, companies tend to adapt their CSR practices according to the implicit or explicit CSR institutional environment. We also control for the time trend including year dummies in the model. Table 7 summarizes our variables.

Table 7: Description of variables

Main variables	Definition	Source
Analytic	LIWC's summary measure of analytic thinking in writing – normalized to 0-1 range.	LIWC
Assertive	Factor 1 derived from a factor analysis (Section 4.3), capturing assertive style in writing – normalized to 0 -1 range.	LIWC (factor analysis)
Defensive	Factor 2 derived from a factor analysis (Section 4.3), capturing defensive style in writing – normalized to 0 -1 range.	LIWC (factor analysis)
INDsen	Dummy variable that takes value 1 if the company operates in sensitive industry (materials, oil&gas, utilities) and 0 otherwise	Bloomberg
ESGcontro	Refinitiv's ESG controversies score – reverse coded and normalized to 0 -1 range.	Refinitiv
Control variables		
EnvSoc	Company-year mean of Refinitiv's environmental and social scores. Scores range from 0 to 100.	Refinitiv
StratDis	Intensity of strategic business case-related content in the CEO letter – normalized to 0 -1 range.	Topic modeling (Appendix 1)
GRI	Dummy variable that takes value 1 if the company reports in accordance with GRI guidelines/standards and 0 otherwise.	Bloomberg

ASSU	Dummy variable that takes value 1 if the companies' SR was subject of independent assurance and 0 otherwise.	Bloomberg
CSRcom	Dummy variable that takes value 1 if the company has a CSR or sustainability committee at the board level and 0 otherwise	Bloomberg
BRDwmn	The percentage of female directors to the total board membership.	Bloomberg
BRDsize	Number of full time directors on the company's board.	Bloomberg
BRDind	The percentage of independent directors to the total board membership.	Bloomberg
CEOdual	Dummy variable that takes value 1 if the CEO is also Chairman of the Board and 0 if the two roles are separate.	Bloomberg
ROA	Return on assets.	Bloomberg
Size	Natural logarithm of total assets.	Bloomberg
UK	Dummy variable that takes value 1 if the company is headquartered in the UK and 0 otherwise	Bloomberg
US	Dummy variable that takes value 1 if the company is headquartered in the US and 0 otherwise	Bloomberg

The hypotheses are tested using random effects models that allow us to estimate the effects of our time-invariant variable of interest (INDsen). The choice of the model was supported by the results of the Mundlak tests (Mundlak, 1978). For assertive, and defensive as explained variables tests showed: Chi-squared(12) = 18.01, Prob > Chi-squared = .113; Chi-squared(13) = 20.91, Prob > Chi-squared = .052, respectively. These results indicate that there is no correlation between time-invariant unobservables and regressors, hence we reject the null hypotheses and conclude that the random effects models apply.

5. Results

5.1. Descriptive statistics

Table 8 provides descriptive statistics to present the overall characteristics of the companies included in the sample. 38% of our observations come from sensitive industries. The mean value for the ESG controversy score is noted at 0.22. While the majority (76%) of the analyzed documents were prepared according to GRI standards/guidelines, merely half of them were subject to independent assurance. With regard to corporate governance practices, the majority of the sample companies have CSR/sustainability committees at the board level, as well as CEO and Chairman of the Board separation. The sample companies reveal variation in terms of board size (with a mean at the level of 11 members), independence (with 75% of independent directors on average) and board gender diversity (with merely 23% of women on boards on average). Performance statistics reveal a mean value for EnvSoc at the level of 66.43 and ROA of 5.51. The sample companies are quite similar though with respect to size, with a mean and standard deviation of company size at the level of 23.44 and 1.44, respectively[16]. Variance inflation factors (VIF) for the dataset consistently fall below the widely accepted threshold (<10) (Hair et al., 2010), indicating the absence of any significant multicollinearity issues.

Table 8: Descriptive statistics (2010-2019)

Variable	**Obs.**[17]	**Mean**	**S.D.**	**Min.**	**Max.**	**VIF**
Analytic	2,139	0.80	0.12	0	1	
Assertive	2,139	0.74	0.15	0	1	
Defensive	2,139	0.35	0.13	0	1	
INDsen	1,805	0.38	0.48	0	1	1.65
ESGcontro	1,805	0.22	0.32	0	1	2.06
EnvSoc	1,805	66.43	16.85	0.13	96.09	1.71
StratDis	2,139	0.26	0.26	0	1	1.99
GRI	1,805	0.76	0.43	0	1	1.28
ASSU	1,805	0.52	0.50	0	1	1.51
CSRcom	1,805	0.89	0.32	0	1	1.18
BRDwmn	1,805	23.04	11.45	0	70	1.32
BRDsize	1,805	11.06	2.63	3	21	1.56
BRDind	1,805	74.96	19.32	0	100	1.99
CEOdual	1,805	0.33	0.47	0	1	1.32
ROA	1,805	5.51	7.17	-39.60	97.60	1.14
Size	1,805	23.44	1.44	19.65	27.03	2.24

Table 9 presents a correlation matrix. We note that membership in sensitive industries (INDsen) is negatively correlated with the two identified narrative styles (being assertive and defensive), while positively correlated with analytical thinking (Analytic), which provides

[16] The sample exhibits a bias towards larger firms, resulting in limited variation in terms of firm size. This observation serves as a supplementary argument to not to use firm size as a proxy for general visibility.
[17] To ensure a more robust factor analysis, we retained the maximum amount of data comprising SR CEO letters: 2,139 observations. Upon including variables from Bloomberg and Refinitiv, the number of observations decrease to 1,805 that we later use for our regression analyses.

initial support for our hypotheses (H1a and H1b). Correlation analysis also indicates that our issue visibility variable (ESGcontro) is positively correlated with assertive and defensive narrative styles. These results provide support for our expectation regarding the prominence of defensive IM (H2a), yet it cannot support our hypothesis on the relationship between issue visibility driven by sustainability controversies and assertive IM (H2b). The correlations among our independent variables do not indicate multicollinearity, revealing the highest correlation to be between assertive and strategic business case content intensity (StratDis) ($r = -.430$).

Table 9: Correlation matrix

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1) Analytic	1.000															
(2) Assertive	-0.438*	1.000														
(3) Defensive	-0.718*	0.000	1.000													
(4) INDsen	0.127*	-0.108*	-0.122*	1.000												
(5) ESGcontro	-0.099*	0.027	0.100*	0.015	1.000											
(6) EnvSoc	-0.010	-0.077*	0.117*	-0.071*	0.238*	1.000										
(7) StratDis	0.106*	-0.430*	0.181*	-0.077*	-0.076*	0.195*	1.000									
(8) GRI	0.092*	-0.085*	-0.043	0.165*	0.031	0.321*	0.172*	1.000								
(9) ASSU	0.147*	-0.221*	0.015	0.028	0.094*	0.401*	0.237*	0.280*	1.000							
(10) CSRcom	-0.001	-0.060	0.035	0.014	0.098*	0.327*	0.057	0.168*	0.237*	1.000						
(11) BRDwmn	-0.116*	0.037	0.141*	-0.198*	-0.017	0.189*	0.114*	0.011	0.125*	0.062*	1.000					
(12) BRDsize	0.033	-0.092*	-0.006	0.042	0.229*	0.280*	0.050	0.155*	0.139*	0.098*	-0.002	1.000				
(13) BRDind	-0.208*	0.324*	0.013	0.061*	0.078*	0.020	-0.450*	0.004	-0.126*	0.062*	0.080*	-0.083*	1.000			
(14) CEOdual	-0.049	0.125*	-0.070*	0.082*	0.112*	0.024	-0.244*	0.029	-0.121*	0.020	-0.004	0.181*	0.277*	1.000		
(15) ROA	-0.122*	0.129*	0.070*	-0.211*	-0.081*	0.063*	-0.043	-0.108*	-0.032	0.030	0.183*	-0.072*	0.058	0.014	1.000	
(16) Size	-0.035	-0.006	0.039	0.152*	0.488*	0.436*	-0.083*	0.220*	0.265*	0.196*	0.007	0.493*	0.227*	0.258*	-0.136*	1.000

* Correlations significant at p<0.01

5.2. Multivariate analysis

The results based on the empirical model are reported in Table 10.

Table 10: Regression results

VARIABLES	(1) Analytic	(2) Assertive	(3) Defensive	(4) Analytic	(5) Assertive	(6) Defensive
INDsen (H1)	0.0252**	-0.0376***	-0.0188*	0.0308***	-0.0371***	-0.0280**
	(0.0098)	(0.0118)	(0.0103)	(0.0105)	(0.0125)	(0.0112)
ESGcontro (H2)	-0.0191**	-0.0056	0.0261**	-0.0085	-0.0046	0.0089
	(0.0095)	(0.0102)	(0.0106)	(0.0118)	(0.0128)	(0.0132)
INDsen x ESGcontro (H3)				-0.0277	-0.00263	0.0453**
				(0.0185)	(0.0202)	(0.0206)
EnvSoc	0.0000	0.0002	0.0001	0.000	0.0002	0.0002
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
StratDis	-0.2380***	-0.1820***	0.4320***	-0.2370***	-0.1810***	0.4310***
	(0.0333)	(0.0367)	(0.0368)	(0.0333)	(0.0368)	(0.0368)
GRI	-0.0097	0.0165*	-0.0026	-0.0100	0.0164*	-0.0021
	(0.0078)	(0.0086)	(0.0087)	(0.0078)	(0.0086)	(0.0087)
ASSUR	0.0153**	-0.0171**	-0.0072	0.0160**	-0.0170**	-0.0083
	(0.0073)	(0.0081)	(0.0081)	(0.0073)	(0.0081)	(0.0081)
CSRcom	0.0069	-0.0190*	-0.0003	0.0076	-0.0189*	-0.0013
	(0.0101)	(0.0111)	(0.0111)	(0.0101)	(0.0112)	(0.0111)
BRDwmn	-0.0005*	0.000	0.0008**	-0.0005*	0.0000	0.0008**
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
BRDsize	0.00230*	-0.0052***	-0.0025	0.0028*	-0.0053***	-0.0023
	(0.0015)	(0.0017)	(0.0017)	(0.0015)	(0.0017)	(0.0016)

BRDind	-0.0004	0.0002	0.0003	-0.0004	0.0002	0.0003
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
CEOdual	0.0028	-0.007	-0.0030	0.0029	-0.0069	-0.0031
	(0.0080)	(0.0089)	(0.0088)	(0.0080)	(0.0089)	(0.0088)
ROA	-0.0004	0.0010**	0.0000	-0.0004	0.0010**	0.000
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Size	-0.0028	0.0014	0.0036	-0.0028	0.0013	0.0037
	(0.0041)	(0.0046)	(0.0043)	(0.0039)	(0.0047)	(0.0043)
UK	-0.0939***	0.0371**	0.0896***	-0.0939***	0.0371**	0.0896***
	(0.0145)	(0.0174)	(0.0153)	(0.0145)	(0.0174)	(0.0153)
US	-0.1120***	0.0980***	0.0852***	-0.1130***	0.0980***	0.0864***
	(0.0144)	(0.0167)	(0.0154)	(0.0144)	(0.0167)	(0.0154)
Year dummies	Included	Included	Included	Included	Included	Included
Constant	0.9600***	0.724***	0.144*	0.961***	0.725***	0.143
	(0.0822)	(0.0966)	(0.0875)	(0.0822)	(0.0967)	(0.0876)
Within R^2	0.04	0.03	0.08	0.05	0.03	0.08
Between R^2	0.22	0.31	0.18	0.22	0.31	0.18
Overall R^2	0.14	0.28	0.14	0.15	0.28	0.14
Observations	1,805	1,805	1,805	1,805	1,805	1,805
Number of companies	345	345	345	345	345	345

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

In the first three models, we test the direct relationship (i.e., main effect) between our dependent and test variables (excluding the interaction term between INDsen and ESGcontro). Model (1) having analytic as the dependent shows that, in line with our expectations (H1a), companies operating in sensitive industries disclose more formal SR CEO letters (i.e., significantly positive coefficient for INDsen). With regard to IM, we find a significantly negative relationship between sensitive industry membership (INDsen) and both the assertive (Model (2)) and defensive (Model (3)) styles. For defensive, however, the association is only marginally significant (i.e., at the 10% significance level). As the level of significance for sensitive industry membership is different for assertive ($b = -0.0376$, $p < 0.01$) and defensive styles ($b = -0.0188$, $p < 0.1$), we assume that it could indicate that the effect of general visibility for assertive is stronger than for defensive. Hence, we test for the significance of the difference between those two coefficients, which proved to be statistically significant at 10% level. Overall, these results provide support for our hypothesis (H1b). We further note a significantly positive relationship between ESG controversies (ESGcontro) and defensive style in SR CEO letters (Model (3)), providing support for H2a. H2b is, however, not supported, as we observe no significant relationship between ESG controversies score and assertive (Model (2)).

In Models (4) up to (6) we then add the interaction term between INDsen and ESGcontro to test the hypotheses regarding the moderating effect (H3a and H3b). The results indicate a significantly positive interaction effect for the defensive IM style (Model (6)), that supports our H3a. The lack of statistically significant coefficient on ESGcontro in Model 6 suggests that it is companies from sensitive industries that are more prone to resort to defensive IM when controversies arise. Figure 1 depicts that when controversies reach a very high level the relationship between sensitive industries and defensive IM becomes slightly positive. We find no statistically significant interaction effect in Model (4), nor Model (5) (H3b was not supported).

Regarding our control variables, our results show the negative effect of strategic business case content intensity on the use of analytic and assertive styles ($b = -.237$ and $b = -.181$, $p < 0.01$) and positive effect on the defensive style ($b = .431$, $p < 0.01$). The positive effect on the defensive style could stem from the increased emphasis on justifications and sensemaking within the strategic content, aligning with the nature of defensive IM. In addition, our results show that independent assurance of SR has a positive effect on the use of formal and analytical language ($b = .016$, $p < 0.05$) while constraining the use of assertive style ($b = -.017$, $p < 0.05$) among with board size ($b = -.005$, $p < 0.01$) and CSR committee ($b = -.018$, $p < 0.1$). Finally, both region dummies (UK and US) move in the same direction in comparison with the EU institutional setting, that is both UK and US negatively affect the use of analytic style ($b = -.094$; $b = -.113$, $p < 0.01$); while positively affect the use of both assertive ($b = .037$; $b = .098$, $p < 0.05$) and defensive style ($b = .090$; $b = .086$, $p < 0.01$). According to Matten and Moon's (2008) perspective, these discrepancies could be associated with unique institutional arrangements at the regional level. Matten and Moon's (2008) implicit-explicit CSR framework suggests that both US- and UK-oriented CSR tends to be more explicit in comparison to the European-style CSR, which provides more room for corporate initiative and therefore both assertive and defensive IM. Lastly, the findings indicate that the proportion of female directors relative to the total board membership has a negative impact on the use of formal and analytical language ($b = .0005$, $p < 0.1$); while exhibiting a positive association with the adoption of a defensive style ($b = .0008$, $p < 0.05$).

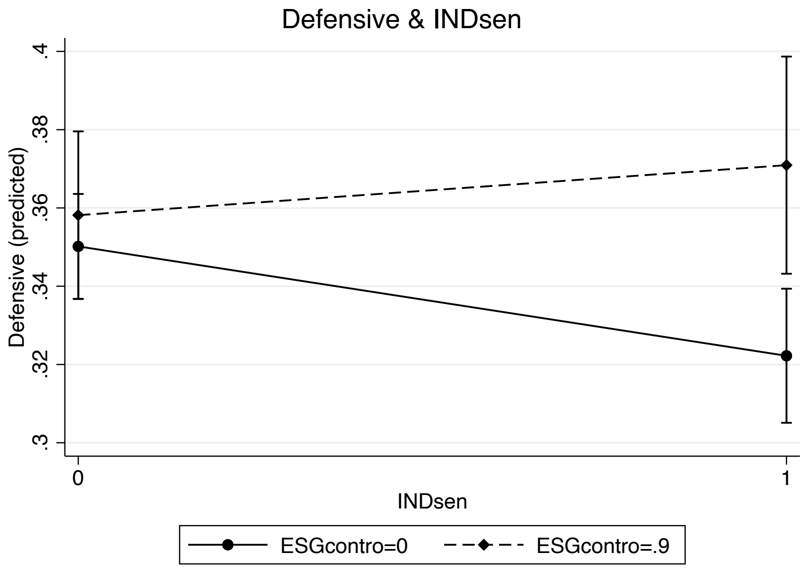

Figure 1: Graphing H3a results

Notes: This figure depicts average predicted values generated from Model 6 for *defensive*. The estimates are made at the 5th percentile and 95th percentile of ESGcontro. Both lines serve the purpose of highlighting the relative differences based on the binary category (INDsen). The solid line depicts estimates for companies with low controversy scores. The dashed line depicts estimates for companies with high controversy scores.

5.3. Robustness and extra analyses

To check whether our results are not biased due to the skewed distribution of ESGcontro[18], we re-estimate our primary models for the subsample of observations with reported ESG controversies (ESGcontro>0). The results of these tests are tabulated in Table 11. Results with regard to the direct effect of sensitive industry membership (INDsen) and ESG controversies (ESGcontro) yield the same sign, yet the statistical significance decreases (and even disappears) for some of the previously observed effects (especially so in Models (1) and (3)). Specifically, the coefficient for INDsen is no longer significant in Model (1) and (3) but when the interaction term is included (Models (4) and (6)), coefficients do attain statistical significance (which is in line with the main analyses). In addition, we also conducted analyses excluding observations from the US, which account for approximately 50% of the dataset. The results (untabulated) indicate a consistent direction of effects but loss in statistical significance for defensive. We associate the decrease in significance levels with the smaller sample size

[18] For the majority of our sample, i.e., 1,048 observations, no controversies have been reported.

used for the robustness check. Nevertheless, the results of the robustness check qualitatively support our primary findings. Winsorizing the continuous variables at the top and bottom tails, at a 1% level, did not yield any significance alterations in our results either.

Table 11: Regression results for the sub-sample ESGcontro>0

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Analytic	Assertive	Defensive	Analytic	Assertive	Defensive
INDsen (H1)	0.0203	-0.0516***	-0.0058	0.0486**	-0.0617***	-0.0350*
	(0.0137)	(0.0141)	(0.0130)	(0.0204)	(0.0205)	(0.0202)
ESGcontro (H2)	-0.0252*	-0.0100	0.0317**	-0.00364	-0.0178	0.00951
	(0.0150)	(0.0147)	(0.0153)	(0.0189)	(0.0186)	(0.0193)
INDsen X ESGcontro (H3)				-0.0585*	0.0211	0.0596*
				(0.0314)	(0.0309)	(0.0320)
EnvSoc	-0.0029	-0.0215	0.0236	-0.00593	-0.0204	0.0265
	(0.0260)	(0.0261)	(0.0257)	(0.0260)	(0.0262)	(0.0256)
StratDis	-0.1990***	-0.2190***	0.4400***	-0.1960***	-0.2200***	0.4360***
	(0.0563)	(0.0561)	(0.0563)	(0.0562)	(0.0561)	(0.0560)
GRI	-0.0172	0.0046	0.0089	-0.0174	0.0047	0.0088
	(0.0125)	(0.0124)	(0.0127)	(0.0125)	(0.0124)	(0.0126)
ASSUR	0.0195	-0.0310***	-0.0048	0.0226*	-0.0321***	-0.0081
	(0.0121)	(0.0120)	(0.0122)	(0.0122)	(0.0121)	(0.0123)
CSRcom	0.0202	0.0107	-0.0356*	0.0235	0.0094	-0.0388*
	(0.0202)	(0.0200)	(0.0206)	(0.0203)	(0.0201)	(0.0206)
BRDwmn	-0.0015***	-0.0001	0.0015***	-0.00150***	-0.0001	0.0015***
	(0.0005)	(0.0005)	(0.0006)	(0.0005)	(0.0005)	(0.0005)
BRDsize	0.0003	-0.0044*	0.0003	0.0000	-0.0042*	0.0006
	(0.0024)	(0.0023)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
BRDind	-0.0009**	0.0004	0.0010**	-0.0009**	0.0003	0.0009**
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
CEOdual	-0.0056	-0.00873	0.0012	-0.0053	-0.0088	0.0008
	(0.0122)	(0.0122)	(0.0121)	(0.0121)	(0.0122)	(0.01200)

ROA	-0.0006	0.0004	0.0003	-0.0006	0.0004	0.0003
	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Size	0.0000	0.0044	-0.0034	-0.0006	0.0046	-0.0028
	(0.0057)	(0.0058)	(0.0055)	(0.00567)	(0.0058)	(0.0054)
UK	-0.0931***	0.0488**	0.1070***	-0.0938***	0.0491**	0.1070***
	(0.0203)	(0.0210)	(0.0195)	(0.0203)	(0.0210)	(0.0193)
US	-0.0725***	0.0777***	0.0634***	-0.0743***	0.0785***	0.0654***
	(0.0219)	(0.0222)	(0.0214)	(0.0219)	(0.0223)	(0.0213)
Year dummies	Included	Included	Included	Included	Included	Included
Constant	0.981***	0.755***	0.158	0.994***	0.750***	0.145
	(0.129)	(0.132)	(0.125)	(0.129)	(0.132)	(0.124)
Within R^2	0.06	0.02	0.12	0.06	0.02	0.12
Between R^2	0.18	0.37	0.23	0.19	0.37	0.23
Overall R^2	0.14	0.3	0.19	0.14	0.31	0.19
Observations	757	757	757	757	757	757
Number of companies	222	222	222	222	222	222

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: Random effects models. N = 757 company-year observations pertaining to 222 companies headquartered in 17 countries.

In addition, we further investigate the effect of strategic business case content intensity in our models due to the significant results and strong coefficient of the variable. We tested the models excluding this control variable and the results hold after the exclusion (untabulated); however, the explanatory power of our models drops substantially. We manually investigated examples of CEO letters with both high and low strategic business case-related content. As highlighted in our main analyses, in CEO letters with high strategic business case-related content defensiveness is more prominent, mainly driven by cognitive processes words, such as insight words (e.g., *affect*, *believe*, etc.), causation (e.g., *because*, *why*, etc.) and discrepancy (e.g., *would, want, need,* etc.), as well as auxiliary verbs that are used for passive sentence construction. The examples of high strategic business case-related CEO letters are provided below:

> *"…We also believe that taking a long-term view is crucial when it comes to dealing with complex sustainability issues. According to the UN, climate change and poverty are two of the most significant challenges of our time and will affect many generations to come. While I have a great deal of respect for the vast complexity of both of these challenges, I also strongly believe that we can make a positive contribution towards facing them… In order for our company to take on these challenges in an effective way I believe that sustainability must be considered as an integral part of our business rather than being treated as an after-thought…. However, it is not possible to achieve great change in isolation. We have over 70 years of experience, but we are entirely dependent on our partnerships with experts from other fields to really drive our sustainability work forward. We therefore fully support the development of the new technology and innovation that is required to create a circular fashion industry… I'm convinced that technical innovations will be the solution to many of the environmental challenges the textile industry is facing and will contribute to a more sustainable consumption…"* - H&M Group (2018, Sustainability Report 2018) (Strategic business case-related content intensity = 0.499).

The manager of the company emphasizes the long-term importance of sustainability, takes an innovative stance and acknowledges the complexity of the difficult nature of sustainability issues like climate change and poverty. The manager cautiously promotes the integration of sustainability as an integral part of the business, yet does not directly mention about steps. The linguistic cues in the CEO letter of H&M Group also shows that the writer is being cautious and indicates a distance from the claims being made. The excessive use of insight words and phrases like "*I have a great deal of respect*" tend to signal a degree of uncertainty and the use of causal connectors like "*however*" or "*therefore*" indicates a tendence for hedging and competing statements. In addition, passive statements can also create a detachment and reducing direct responsibility commitment to the claims as in statement: "*however, it is not possible to achieve great change in isolation. We have over 70 years of experience, but we are entirely dependent on our partnerships with experts from other fields to really drive our sustainability work forward".*

> *"...I believe that only a sustainable company, with sustainable growth, is able to deliver sustainable solutions. To demonstrate the seriousness of our vision, the Volvo Group has revised our CSR and sustainability strategy and is incorporating it into our daily work. I feel that we have a method that is clearly connected to our business model and our strategies... I believe that a responsible company has better prerequisites for becoming a credible business partner. To meet society's needs for sustainable transport solutions, we must primarily cooperate with our customers but also with other players in the industry and society. To operate in many of our new markets, it is also a prerequisite to contribute to social development. Another example that I would like to highlight is the Volvo step, our one-year training course for unemployed young people in Sweden, which started in 2012 with 4,000 applicants for the 400 first positions. The Volvo step is an investment in securing our supply of expertise, while contributing something to reducing unemployment in young people in Sweden. One lesson learned from this three-year project is that we in the industry must listen more to young people..."* - Volvo Group (2012, Sustainability Report 2012: Strategic Approach) (Strategic business case-related content intensity = 0.482).

The passage from the CEO letter of Volvo Group's SR also emphasizes the interconnectedness between sustainability and responsible business practices taking a more defensive stance through connecting sustainability methods directly to the business model and strategies; meanwhile also highlighting co-work with other business partners. Such content is reflected to the linguistic style with the use of phrases, such as "*I believe*" and "*I feel*", and connectors, such as "*but also*", which is used to introduce contrasting viewpoints. Although there are no direct passive statements in the text, the defensiveness emerges from the frequent reliance on personal beliefs and examples and discrepancy words such as "*would*". In contrast, companies that have lower score are expected to be more analytic and assertive.

> *"...As one of the first independent power producers in the U.S. and a multinational electric power company, AES has been at the forefront of bringing innovation to electricity generation and distribution since its founding. AES has a thirty-year track*

> *record of successfully meeting local challenges by bringing our global knowledge and innovation to bear to create highly efficient infrastructure solutions. The diverse mix of our electric generation portfolio and deep expertise in industry technologies provide AES the strength and flexibility to maximize plant efficiency and availability. We deliver reliable, affordable electricity and at the same time we seek to minimize environmental impacts within the technological, economic, and market constraints that we face. In Chile, our Angamos plant exemplifies the way AES brings innovation to unique local challenges. Angamos is a coal-hybrid facility that uses the first-of-its-kind sea water cooling tower in South America. It also incorporates 40MWs of resource equivalent battery storage. The facility received "Power Plant of the Year" by Power Magazine and also won our industry's most prestigious award, the Edison Electric Institute's (EEI) "International Edison Award." At Changuinola, Panama, we brought a 223 MW hydroelectric complex online. Constructing this innovative dam, hydroelectric plant and reservoir required very careful assessments, and complex civil, environmental and social engineering and planning. We are proud of the track record of expert innovation that allows us to undertake and successfully implement projects such as Changuinola..." – The AES Corporation (2012, AES Annual Sustainability Report)* (Strategic business case-related content intensity = 0.0001).

The provided example by the AES Corporation's SR CEO letter has a very low strategic business case-related content intensity, showing how fewer narrative elements are used, while presenting sustainability-related facts, achievements and technical details about the company's power generation projects in a very confident way. The text employs a more formal and analytic stance, without engaging the audience due to lack of use of personal pronouns. The language of such CEO letters is also expected to be written more for informational goals rather than persuasive goals. The assertive style seems to be also found with CEO letters that have less strategic business case-related content:

> *"In 2014, GM took important steps on its journey to become the most valued automotive company. We made significant strides in the face of an extraordinary set of challenges that we took head on, and we are a stronger company today because of what we learned and the progress we have made. our core operating results in 2014 reflect what's possible as we work to deliver more sustainable value for our company as well as our customers and communities. At GM, strengthening our company while building stronger communities and a better world through improved mobility defines our approach to sustainability. Personal mobility means freedom; it means economic advancement; it means connected families and communities; it also means a world of safer and smarter vehicles – cars, trucks and crossovers that use less fuel; that have less environmental impact; and that are programmed to help drivers avoid accidents and reduce congestion. By far, people are the most important drivers of these efforts, and every other initiative at our company. Building a winning culture is a priority and a key takeaway from the issues of 2014, which underscored our need to change behaviors within GM. this winning culture demands candor, accountability and an unwavering focus on customers…"* - General Motors (2014, GM Sustainability Report) (Strategic business case-related content intensity = 0.002).

The paragraph provided by General Motors focuses more on lessons learned and opportunities for future growth, reflecting a more people-centric approach. The content of the CEO letter emphasizes progress and company's strengths, rather than focusing on taking a defensive approach. The text utilizes personal pronouns widely, such as "*we*" and "*our*" emphasizing collective involvement, which aligns with the positive correlation between assertive and personal pronouns. In addition, the passage highlights achievements, sustainability-related goals and progress made by GM, reflecting a positive and achievement-oriented tone.

6. Discussion and conclusion

Our paper investigates the effect of general and issue visibility on the use of rhetorical IM in SR CEO letters. Rhetorical IM can be a powerful tool for managers to shape public perceptions of their companies' ESG track record. By carefully selecting and framing information, companies can present themselves in a favorable light and downplay any unfavorable aspects of their operations. SR, therefore, can be used as a legitimacy tool to communicate with stakeholders about their needs and demands to attain legitimacy even without any substantial corporate action (Bansal & Kistruck, 2006; De Villiers & Van Staden, 2006). While prior research presents several kinds of rhetorical legitimation strategies and IM techniques in SR, research on the factors influencing the use of such opportunistic behavior remains deficient. As prior research in the field of social psychology underscores the significance of publicity as a fundamental driver of IM (Leary & Kowalski, 1990), we argue that visibility notions play a significant role in IM in SR. Prior literature offers contradictory findings about the impact of visibility on the use of IM: some argue that greater visibility prevents the use of IM, due to increased scrutiny and institutional constraints, while others claim that greater visibility promotes IM, mainly due to legitimacy concerns.

The findings of our study make several contributions to the current literature. First, drawing on socio-political theories, we posit that distinct motivations associated with legitimacy, namely maintaining legitimacy or repairing legitimacy, are activated through general visibility and issue visibility. Our paper shows that general visibility related to sensitive industry membership, where scrutiny for IM is clearly higher, leads to more formal language. Companies operating in sensitive industries are less likely to resort to IM compared to companies operating in non-sensitive industries. While the results are in line with our expectations for both assertive and defensive types of rhetorical IM, we find that the impact of general visibility on the use of the assertive style is stronger. The level of scrutiny and activist pressure on companies operating in non-sensitive industries is significantly lower (Marquis et al., 2016). As the standards and regulations are also less stringent for these companies, managers tend to use assertive style to describe organizational outcomes and prospects to trigger more favorable impressions (Henry, 2008; Tan et al., 2014). Our findings indicate that sensitive industry members tend to disclose SR in response to accountability and transparency concerns of stakeholders to provide a credible image in response to scrutiny. Drawing from our interpretation of outcomes, it can be deduced that the regulatory controls, governance structures and reporting frameworks for sensitive industries serve their intended purposes; while for the non-sensitive industries managerial leeway for IM persists, attributed to the comparatively limited presence of stringent control and governance mechanisms. This shows that within sensitive industries, companies tend to follow the same regulations, maintain their legitimacy and show mimetic behavior about accountable and transparent reporting. While these results are similar to the findings of previous studies, such as Marquis et al. (2016), Ruiz-Blanco et al. (2021), which *consistently indicate that not only do firms with a higher pollution propensity disclose more environmental information; they also rely on disclosures that the GRI views as* *inherently more objective and verifiable* (Clarkson et al., 2011, p.27), we further argue that a distraction in legitimacy may change this reporting behavior.

Second, as O'Donovan (2002) argues, the legitimation techniques chosen depends on whether the company is trying to maintain the current level of legitimacy or repair and gain the damaged and threatened legitimacy. Controversies, as such, disrupt the established state of legitimacy. Corporate controversies rapidly attract media attention through news and social media and create/increase the discrepancy between the desired and current image of the company (DasGupta, 2021; Del Giudice & Rigamonti, 2020; Elsbach & Sutton, 1992). This rapid break in the legitimacy status-quo through media channels leads managers to become more prone to use IM strategies. This allows us to extend prior literature on IM in SR by investigating the effect of controversies. Specifically, our findings show that in the face of a legitimacy threat, managers' disclosure strategies shift towards justifications and sense-giving mechanisms. Despite the stringent regulations and governance in sensitive industries, the effect of controversies is significant and positive on the use of defensive IM strategies, while not significant for non-sensitive industries. As such, the relationship between general visibility and IM is context-dependent, which can vary due to controversies. Restoring legitimacy from the detrimental impact of unethical or irresponsible behavior and reputational loss may be more important for sensitive and "dirty" industries, possibly due to their higher business exposure, in line with legitimacy theory. Legitimacy theory is considered one of the most common theories in the sustainability disclosure literature, despite receiving persistent critique from numerous scholars (Patten, 2019). Yet, the theory seems to be still prominent, and our study further contributes to a deeper understanding of it through showing the positive relationship between controversies and defensive IM. At the same time, contrary to our expectations, we cannot find empirical evidence to prove the negative relationship between sustainability controversies and assertive IM. Previous studies highlight a prevailing inclination towards the usage of assertive IM in SR (e.g., Barkemeyer et al., 2014; Caliskan et al., 2021), but we argue that in SR, managers allow themselves to be assertive only when they are not afraid of being accused of greenwashing or window-dressing. This may indicate that motives for being both assertive and defensive may be different in SR than in financial reports.

Finally, in our analyses, we show that thematic content is a significant variable that affects managers' use of language (e.g., linguistic style), which has been largely overlooked in prior literature. In all our models, we show that strategic business case content intensity is a significant explanatory variable. More specifically, our results show that strategic business case related content leads to a more defensive language. This defensiveness may relate to the explanations and justifications provided to readers interested in how sustainability is integrated into business strategies. The results regarding the countries show opposite directions for the use of IM, and types of IM, which may relate to the country-level institutional settings; as well as the sustainability-related knowledge and culture among the societies. Specifically, our results show that both the UK and US (with Europe being the benchmark) are positively associated with IM style (both assertive and defensive) and negatively associated with formal language (analytic). As Matten and Moon (2008) argue, such differences may relate to distinct country-level institutional settings. Further, they proposed the implicit-explicit CSR framework to compare and contrast country-level institutional settings. The implicit and explicit nature of CSR is rooted in the national business systems, where the norms, incentives and rules shaping CSR are defined by the government and markets within the prevailing institutional framework (Matten & Moon, 2008). Companies, in turn, adapt their CSR practices in accordance with these established parameters. According to the framework, both the US- and UK-style CSR is more explicit compared to the European-style CSR (LaGore et al., 2020; Matten & Moon, 2008). As explicit CSR offers much more room for corporate initiative, we observe in our results the positive effect of it in constructing both assertive and defensive IM strategies.

In addition, to the best of our knowledge, our paper is one of the first papers that quantitatively analyses the effects of industry and controversies on the use of rhetorical IM in SR. Previous studies have primarily concentrated on the thematic content in SR, such as selective disclosure, legitimization strategies and the recognized performance-disclosure gap (e.g., García-Sánchez et al., 2022; Macellari et al., 2021; Marquis et al., 2016; Roszkowska-Menkes et al., 2024; Ruiz-Blanco et al., 2021), but there has been insufficient understanding towards the style-related IM, which includes the rhetorical strategies employed by managers. This way, we propose a more refined approach to focus on linguistic style in SR CEO letters and extend the existing limited body of research, which predominantly depends on qualitative methods involving small sample sizes. Another strength of our study is that we combine formal language (no IM) with assertive and defensive IM strategies using a multi-industry and -country dataset, Thomson Reuters – Eikon and Bloomberg. This provides a more complete picture of the array of rhetorical strategies utilized by managers in SR CEO letters. Additionally, we explicitly differentiate between narrative and non-narrative sections of SR as suggested by Fuoli (2018), recognizing the unique role narratives play in shaping perceptions and creating a storytelling atmosphere. Prior research on IM in SR has not made this differentiation clearly, nor focused on discourse directly by managers. For this reason, our framework also incorporates novel machine learning and NLP techniques in understanding IM. As we show the significant effect of controversies, further research may investigate rhetorical IM at a topical level, as well as to understand which pillar of sustainability is subject to IM.

Overall, our study contributes to the literature by empirically showing general visibility and issue visibility triggers different legitimacy motives. Investigating the rhetorical IM in SR empirically and showing where we observe the formal use of language and IM may enlighten researchers and policymakers regarding the transparency of sustainability-related information and accountability in SR. In addition, our research helps investors and analysts assess companies and their corporate communication strategies, which can eventually shape socially responsible investor behaviors and contribute to developing a more effective governance and regulation system in the realm of SR, eventually contributing to transparency and accountability purpose of SR.

A major limitation of our study is the availability of SR of companies. The reliance on databases is insufficient, necessitating a considerable amount of time in manual data collection process. This resulted in a drop in expected SR for our analysis. Another limitation of the study is the primarily focus on large companies. As highlighted by Patten (2002), company size is a factor that also influences general visibility and SR disclosure quality. However, in our models, size remained insignificant. This insignificance may be attributed to our emphasis on large companies within our database. We recognize that our analysis does not include controls for CEO-specific motivations, which have been shown to influence CSR disclosure practices (Lassoued & Khanchel, 2022), nor does it incorporate variables related to information demand (such as analyst coverage and ownership structure) due to data limitations. Finally, our variable selection for factor analysis includes summary and composite variables, that include various linguistic styles. For example, our cognitive processes variable is composed of insight words, causation words, discrepancy words, tentative words, certitude words and differentiation words. This approach restricts us from focusing on a specific linguistic style yet provides us with a general overview of rhetorical IM patterns when used in factor analysis. Notwithstanding these limitations, we suggest that further research may explore and shed light on the factors and policies affecting the accountability of sustainability-related information, as well as the effect of other corporate governance factors, which may restrict managers' use of SR as a symbolic legitimacy tool. We believe that AI and machine learning tools are also essential for future research to identify both semantic and thematic language control. Continued efforts are needed to make SR more accountable and transparent as financial reports. In final words, we would like to refer to Patten (2019): "*Legitimacy-based research can help move CSR disclosure at least closer to being a tool of accountability, as opposed to a tool for legitimation".*

Reference List

Aerts, W., Cheng, P., & Tarca, A. (2013). Management's Earnings Justification and Earnings Management under Different Institutional Regimes. *Corporate Governance: An International Review, 21*(1), 93-115. https://doi.org/10.1111/corg.12014

Aerts, W., & Cormier, D. (2009). Media legitimacy and corporate environmental communication. *Accounting, Organizations and Society, 34*(1), 1-27. https://doi.org/10.1016/j.aos.2008.02.005

Aerts, W., & Tarca, A. (2010). Financial performance explanations and institutional setting. *Accounting and Business Research, 40*(5), 421-450. https://doi.org/10.1080/00014788.2010.9995322

Aerts, W., & Yan, B. (2017). Rhetorical impression management in the letter to shareholders and institutional setting. *Accounting, Auditing & Accountability Journal, 30*(2), 404-432. https://doi.org/10.1108/aaaj-01-2015-1916

Akstinaite, V., Robinson, G., & Sadler-Smith, E. (2020). Linguistic Markers of CEO Hubris. *Journal of Business Ethics, 167*, 687-705. https://doi.org/10.1007/s10551-019-04183-y

Alonso-Almeida, M. d. M., Marimon, F., Casani, F., & Rodriguez-Pomeda, J. (2015). Diffusion of sustainability reporting in universities: current situation and future perspectives. *Journal of Cleaner Production, 106*, 144-154. https://doi.org/10.1016/j.jclepro.2014.02.008

Amernic, J. H., & Craig, R. J. (2007). Guidelines for CEO-speak: editing the language of corporate leadership. *Strategy & Leadership, 35*(3), 25-31. https://doi.org/10.1108/10878570710745802

Amran, A., Lee, S. P., & Devi, S. S. (2014). The Influence of Governance Structure and Strategic Corporate Social Responsibility Toward Sustainability Reporting Quality. *Business Strategy and the Environment, 23*(4), 217-235. https://doi.org/10.1002/bse.1767

Aouadi, A., & Marsat, S. (2018). Do ESG Controversies Matter for Firm Value? Evidence from International Data. *Journal of Business Ethics, 151*(4), 1027-1047. https://doi.org/10.1007/s10551-016-3213-8

Bagnoli, M., & Watts, S. G. (2017). Voluntary Assurance of Voluntary CSR Disclosure. *Journal of Economics & Management Strategy, 26*(1), 205-230. https://doi.org/10.1111/jems.12171

Bansal, P., & Kistruck, G. (2006). Seeing Is (Not) Believing: Managing the Impressions of the Firm's Commitment to the Natural Environment. *Journal of Business Ethics, 67*(2), 165-180. https://doi.org/10.1007/s10551-006-9021-9

Barkemeyer, R., Comyns, B., Figge, F., & Napolitano, G. (2014). CEO statements in sustainability reports: Substantive information or background noise? *Accounting Forum, 38*(4), 241-257. https://doi.org/10.1016/j.accfor.2014.07.002

Ben-Amar, W., Chang, M., & McIlkenny, P. (2017). Board Gender Diversity and Corporate Response to Sustainability Initiatives: Evidence from the Carbon Disclosure Project. *Journal of Business Ethics, 142*(2), 369-383. https://doi.org/10.1007/s10551-015-2759-1

Boiral, O. (2013). Sustainability reports as simulacra? A counter-account of A and A+ GRI reports. *Accounting, Auditing & Accountability Journal, 26*(7), 1036-1071. https://doi.org/10.1108/aaaj-04-2012-00998

Boiral, O. (2016). Accounting for the Unaccountable: Biodiversity Reporting and Impression Management. *Journal of Business Ethics, 135*(4), 751-768. https://doi.org/10.1007/s10551-014-2497-9

Boiral, O., Brotherton, M.-C., & Talbot, D. (2020). Building trust in the fabric of sustainability ratings: An impression management perspective. *Journal of Cleaner Production, 260*. https://doi.org/10.1016/j.jclepro.2020.120942

Bolino, M. C., Kacmar, K. M., Turnley, W. H., & Gilstrap, J. B. (2008). A Multi-Level Review of Impression Management Motives and Behaviors. *Journal of Management, 34*(6), 1080-1109. https://doi.org/10.1177/0149206308324325

Bowen, F. E. (2000). Environmental visibility: a trigger of green organizational response? *Business Strategy and the Environment, 9*(2), 92-107. https://doi.org/10.1002/(sici)1099-0836(200003/04)9:2<92::Aid-bse230>3.0.Co;2-x

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). *The Development and Psychmetric Propoerties of LIWC-22*. The University of Texas at Austin.

Bozzolan, S., Cho, C. H., & Michelon, G. (2015). Impression Management and Organizational Audiences: The Fiat Group Case. *Journal of Business Ethics, 126*(1), 143-165. https://doi.org/10.1007/s10551-013-1991-9

Braam, G., Uit de Weerd, L., Hauck, M., & Huijbregts, M. A. J. (2016). Determinants of corporate environmental reporting: the importance of environmental performance and assurance. *Journal of Cleaner Production, 129*, 724-734. https://doi.org/10.1016/j.jclepro.2016.03.039

Bromley, P., & Powell, W. W. (2012). From Smoke and Mirrors to Walking the Talk: Decoupling in the Contemporary World. *The Academy of Management Annals, 6*(1), 483-530. https://doi.org/10.1080/19416520.2012.684462

Brownlow, S., Fogleman, E. L., & Hirsch, S. (2020). How Self-Reflection Influences Use of Cognitive and Analytical Language. *Psychology and Cognitive Sciences – Open Journal, 6*(1), 11-14. https://doi.org/10.17140/pcsoj-6-154

Caliskan, A. O., Esen, E., & Barkemeyer, R. (2021). Impression management tactics in the CEO statements of Turkish sustainability reports. *Business Ethics, the Environment & Responsibility, 30*(4), 485-506. https://doi.org/10.1111/beer.12374

Chelli, M., Durocher, S., & Fortin, A. (2018). Normativity in Environmental Reporting: A Comparison of Three Regimes. *Journal of Business Ethics, 149*(2), 285-311. https://doi.org/10.1007/s10551-016-3128-4

Cho, C. H., & Patten, D. M. (2007). The role of environmental disclosures as tools of legitimacy: A research note. *Accounting, Organizations and Society, 32*(7-8), 639-647. https://doi.org/10.1016/j.aos.2006.09.009

Cho, C. H., Patten, D. M., & Michelon, G. (2012). Impression Management in Sustainability Reports: An Empirical Investigation of the Use of Graphs. *Accounting and the Public Interest, 12*(1), 16-37. https://doi.org/10.2308/apin-10249

Cho, C. H., Roberts, R. W., & Patten, D. M. (2010). The language of US corporate environmental disclosure. *Accounting, Organizations and Society, 35*(4), 431-443. https://doi.org/10.1016/j.aos.2009.10.002

Clarkson, P. M., Li, Y., Richardson, G. D., & Vasvari, F. P. (2008). Revisiting the relation between environmental performance and environmental disclosure: An empirical analysis. *Accounting, Organizations and Society, 33*(4-5), 303-327. https://doi.org/10.1016/j.aos.2007.05.003

Clarkson, P. M., Li, Y., Richardson, G. D., & Vasvari, F. P. (2011). Does it really pay to be green? Determinants and consequences of proactive environmental strategies. *Journal of Accounting and Public Policy, 30*(2), 122-144. https://doi.org/10.1016/j.jaccpubpol.2010.09.013

Cong, Y., Freedman, M., & Park, J. D. (2014). Tone at the top: CEO environmental rhetoric and environmental performance. *Advances in Accounting, 30*(2), 322-327. https://doi.org/10.1016/j.adiac.2014.09.007

Crilly, D., Zollo, M., & Hansen, M. T. (2012). Faking It or Muddling Through? Understanding Decoupling in Response to Stakeholder Pressures. *Academy of Management Journal, 55*(6), 1429-1448. https://doi.org/10.5465/amj.2010.0697

DasGupta, R. (2021). Financial performance shortfall, ESG controversies, and ESG performance: Evidence from firms around the world. *Finance Research Letters*. https://doi.org/10.1016/j.frl.2021.102487

Dawkins, C., & Fraas, J. W. (2010). Coming Clean: The Impact of Environmental Performance and Visibility on Corporate Climate Change Disclosure. *Journal of Business Ethics, 100*(2), 303-322. https://doi.org/10.1007/s10551-010-0681-0

De Villiers, C., & Van Staden, C. J. (2006). Can less environmental disclosure have a legitimising effect? Evidence from Africa. *Accounting, Organizations and Society, 31*(8), 763-781. https://doi.org/10.1016/j.aos.2006.03.001

Del Giudice, A., & Rigamonti, S. (2020). Does Audit Improve the Quality of ESG Scores? Evidence from Corporate Misconduct. *Sustainability, 12*(14). https://doi.org/10.3390/su12145670

Delmas, M., A., & Burbano, V., Cuerel. (2011). The Drivers of Greenwashing. *California Management Review, 54*(1), 64-87. https://doi.org/10.1525/cmr.2011.54.1.64

DiMaggio, P. J., & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review, 48*(2), 147-160. https://doi.org/10.2307/2095101

Diouf, D., & Boiral, O. (2017). The quality of sustainability reports and impression management. *Accounting, Auditing & Accountability Journal, 30*(3), 643-667. https://doi.org/10.1108/aaaj-04-2015-2044

Du, S., & Yu, K. (2020). Do Corporate Social Responsibility Reports Convey Value Relevant Information? Evidence from Report Readability and Tone. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-020-04496-3

Dye, R. A. (1985). Disclosure of Nonproprietary Information. *Journal of Accounting Research, 23*(1), 123-145. https://doi.org/10.2307/2490910

Elsbach, K., D. (1994). Managing Organizational Legitimacy in the California Cattle Industry: The Construction and Effectiveness of Verbal Accounts. *Administrative Science Quarterly, 39*(1), 57-88. https://doi.org/10.2307/2393494

Elsbach, K. D., & Sutton, R. I. (1992). Acquiring Organizational Legitimacy Through Illegitimate Actions: A Marriage of Institutional and Impression Management Theories. *Academy of Management Journal, 35*(4), 699-738. https://doi.org/10.5465/256313

Fabrizio, K. R., & Kim, E.-H. (2019). Reluctant Disclosure and Transparency: Evidence from Environmental Disclosures. *Organization Science, 30*(6), 1207-1231. https://doi.org/10.1287/orsc.2019.1298
Fehre, K., & Weber, F. (2016). Challenging corporate commitment to CSR. *Management Research Review, 39*(11), 1410-1430. https://doi.org/10.1108/mrr-03-2015-0063
Fiandrino, S., & Tonelli, A. (2021). A Text-Mining Analysis on the Review of the Non-Financial Reporting Directive: Bringing Value Creation for Stakeholders into Accounting. *Sustainability, 13*(2). https://doi.org/10.3390/su13020763
Frias-Aceituno, J. V., Rodríguez-Ariza, L., & Garcia-Sánchez, I. M. (2014). Explanatory Factors of Integrated Sustainability and Financial Reporting. *Business Strategy and the Environment, 23*(1), 56-72. https://doi.org/10.1002/bse.1765
Fuoli, M. (2018). Building a Trustworthy Corporate Identity: A Corpus-Based Analysis of Stance in Annual and Corporate Social Responsibility Reports. *Applied Linguistics, 39*(6), 846-885. https://doi.org/10.1093/applin/amw058
García-Meca, E., & Martínez-Ferrero, J. (2021). Is SDG reporting substantial or symbolic? An examination of controversial and environmentally sensitive industries. *Journal of Cleaner Production, 298*. https://doi.org/10.1016/j.jclepro.2021.126781
García-Sánchez, I.-M., Hussain, N., Aibar-Guzmán, C., & Aibar-Guzmán, B. (2022). Assurance of corporate social responsibility reports: Does it reduce decoupling practices? *Business Ethics, the Environment & Responsibility, 31*(1), 118-138. https://doi.org/10.1111/beer.12394
Godfrey, J., Mather, P., & Ramsay, A. (2003). Earnings and Impression Management in Financial Reports: The Case of CEO Changes. *Abacus, 39*(1), 95-123. https://doi.org/10.1111/1467-6281.00122
Greenwood, R., Raynard, M., Kodeih, F., Micelotta, E. R., & Lounsbury, M. (2011). Institutional Complexity and Organizational Responses. *The Academy of Management Annals, 5*(1), 317-371. https://doi.org/10.1080/19416520.2011.590299
Hahn, R., & Kühnen, M. (2013). Determinants of sustainability reporting: a review of results, trends, theory, and opportunities in an expanding field of research. *Journal of Cleaner Production, 59*, 5-21. https://doi.org/10.1016/j.jclepro.2013.07.005
Hahn, R., & Lülfs, R. (2014). Legitimizing Negative Aspects in GRI-Oriented Sustainability Reporting: A Qualitative Analysis of Corporate Disclosure Strategies. *Journal of Business Ethics, 123*(3), 401-420. https://doi.org/10.1007/s10551-013-1801-4
Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate Data Analysis, Englewood Cliffs, NJ: Prentice Hall.
Helfaya, A., & Moussa, T. (2017). Do Board's Corporate Social Responsibility Strategy and Orientation Influence Environmental Sustainability Disclosure? UK Evidence. *Business Strategy and the Environment, 26*(8), 1061-1077. https://doi.org/10.1002/bse.1960
Henry, E. (2008). Are Investors Influenced By How Earnings Press Releases Are Written? *Journal of Business Communication, 45*(4), 363-407. https://doi.org/10.1177/0021943608319388
Hooghiemstra, R. (2000). Corporate Communication and Impression Management – New Perspectives Why Companies Engage in Corporate Social Reporting. *Journal of Business Ethics, 27*, 55-68. https://doi.org/10.1023/A:1006400707757
Hyland, K. (1998). Exploring Corporate Rhetoric: Metadiscourse in the CEO's Letter. *The Journal of Business Communication, 35*(2), 224245. https://doi.org/10.1177/002194369803500203
Hyland, K. (2005). *Metadiscourse : exploring interaction in writing*. Continuum.

Im, J., Kim, H., & Miao, L. (2021, Jan). CEO letters: Hospitality corporate narratives during the COVID-19 pandemic. *Int J Hosp Manag, 92*, 102701. https://doi.org/10.1016/j.ijhm.2020.102701

Ioannou, I., & Serafeim, G. (2012). What drives corporate social performance? The role of nation-level institutions. *Journal of International Business Studies, 43*(9), 834-864. https://doi.org/10.1057/jibs.2012.26

Jackson, G., & Apostolakou, A. (2010). Corporate Social Responsibility in Western Europe: An Institutional Mirror or Substitute? *Journal of Business Ethics, 94*(3), 371-394. https://doi.org/10.1007/s10551-009-0269-8

Jonäll, K., & Rimmel, G. (2010). CEO letters as legitimacy builders: coupling text to numbers. *Journal of Human Resource Costing & Accounting, 14*(4), 307-328. https://doi.org/10.1108/14013381011105975

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology, 33*(2), 125-143. https://doi.org/10.1177/0261927x13502654

Kibler, E., Mandl, C., Farny, S., & Salmivaara, V. (2021). Post-failure impression management: A typology of entrepreneurs' public narratives after business closure. *Human Relations, 74*(2), 286-318. https://doi.org/10.1177/0018726719899465

Kim, P. H., Ferrin, D. L., Cooper, C. D., & Dirks, K. T. (2004, Feb). Removing the shadow of suspicion: the effects of apology versus denial for repairing competence- versus integrity-based trust violations. *J Appl Psychol, 89*(1), 104-118. https://doi.org/10.1037/0021-9010.89.1.104

KPMG. (2020). *The time has come* (The KPMG Survey of Sustainability Reporting 2020, Issue.

Kuruppu, S. C., Milne, M. J., & Tilt, C. A. (2019). Gaining, maintaining and repairing organisational legitimacy. *Accounting, Auditing & Accountability Journal, 32*(7), 2062-2087. https://doi.org/10.1108/aaaj-03-2013-1282

LaGore, W., Mahoney, L., & Thorne, L. (2020). An implicit-explicit examination of differences in CSR practices between the USA and Europe. *Society and Business Review, 15*(3), 165-187. https://doi.org/10.1108/sbr-10-2019-0129

Larrinaga-González, C. (2007). Sustainability reporting: insights from neoinstitutional theory. In *Sustainability Accounting and Accountability* (1st Edition ed.). Routledge.

Lassoued, N., & Khanchel, I. (2022). Voluntary CSR disclosure and CEO narcissism: the moderating role of CEO duality and board gender diversity. *Review of Managerial Science, 17*(3), 1075-1123. https://doi.org/10.1007/s11846-022-00555-3

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin, 107*(1), 34–47. https://doi.org/10.1037/0033-2909.107.1.34

Lyon, T. P., & Maxwell, J. W. (2011). Greenwash: Corporate Environmental Disclosure under Threat of Audit. *Journal of Economics & Management Strategy, 20*(1), 3-41. https://doi.org/10.1111/j.1530-9134.2010.00282.x

Macellari, M., Yuriev, A., Testa, F., & Boiral, O. (2021). Exploring bluewashing practices of alleged sustainability leaders through a counter-accounting analysis. *Environmental Impact Assessment Review, 86*. https://doi.org/10.1016/j.eiar.2020.106489

Marais, M. (2012). CEO rhetorical strategies for corporate social responsibility (CSR). *Society and Business Review, 7*(3), 223-243. https://doi.org/10.1108/17465681211271314

Marquis, C., Toffel, M. W., & Zhou, Y. (2016). Scrutiny, Norms, and Selective Disclosure: A Global Study of Greenwashing. *Organization Science, 27*(2), 483-504. https://doi.org/10.1287/orsc.2015.1039

Matten, D., & Moon, J. (2008). “Implicit” and “Explicit” CSR: A Conceptual Framework for a Comparative Understanding of Corporate Social Responsibility. *Academy of Management Journal, 33*(2), 404-424.

Melloni, G., Stacchezzini, R., & Lai, A. (2016). The tone of business model disclosure: an impression management analysis of the integrated reports. *Journal of Management & Governance, 20*(2), 295-320. https://doi.org/10.1007/s10997-015-9319-z

Meng, X. H., Zeng, S. X., Shi, J. J., Qi, G. Y., & Zhang, Z. B. (2014, Dec 1). The relationship between corporate environmental performance and environmental disclosure: an empirical study in China. *J Environ Manage, 145*, 357-367. https://doi.org/10.1016/j.jenvman.2014.07.009

Merkl-Davies, D. M., & Brennan, N. M. (2017). Theoretical framework of external accounting communication: Research perspectives, traditions, and theories. *Accounting, Auditing & Accountability Journal, 30*(2), 433-469. https://doi.org/10.1108/AAAJ-04-2015-2039

Merkl-Davies, D. M., & Brennan, N. M. (2007). Discretionary disclosure strategies in corporate narratives : incremental information or impression management? *Journal of Accounting Literature, 26*(116-196).

Meyer, J. W., & Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology, 83*(2). https://doi.org/10.1086/226550

Michelon, G., & Parbonetti, A. (2012). The effect of corporate governance on sustainability disclosure. *Journal of Management & Governance, 16*(3), 477-509. https://doi.org/10.1007/s10997-010-9160-3

Michelon, G., Boesso, G., & Kumar, K. (2013). Examining the Link between Strategic Corporate Social Responsibility and Company Performance: An Analysis of the Best Corporate Citizens. *Corporate Social Responsibility and Environmental Management, 20,* 81-94. https://doi.org/10.1002/csr.1278

Michelon, G., Pilonato, S., & Ricceri, F. (2015). CSR reporting practices and the quality of disclosure: An empirical analysis. *Critical Perspectives on Accounting, 33*, 59-78. https://doi.org/10.1016/j.cpa.2014.10.003

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica, 46*(1), 69-85. https://doi.org/10.2307/1913646

Muslu, V., Mutlu, S., Radhakrishnan, S., & Tsang, A. (2019). Corporate Social Responsibility Report Narratives and Analyst Forecast Accuracy. *Journal of Business Ethics, 154*(4), 1119-1142. https://doi.org/10.1007/s10551-016-3429-7

Na, H. J., Lee, K. C., Choi, S. U., & Kim, S. T. (2020). Exploring CEO Messages in Sustainability Management Reports: Applying Sentiment Mining and Sustainability Balanced Scorecard Methods. *Sustainability, 12*(2). https://doi.org/10.3390/su12020590

Nilipour, A., de Silva, T.-A., & Li, X. (2020). The Readability of Sustainability Reporting in New Zealand over time. *Australasian Accounting, Business and Finance Journal, 14*(3), 86-107. https://doi.org/10.14453/aabfj.v14i3.7

O’Donovan, G. (2002). Environmental disclosures in the annual report. *Accounting, Auditing & Accountability Journal, 15*(3), 344-371. https://doi.org/10.1108/09513570210435870

Ogden, S., & Clarke, J. (2005). Customer disclosures, impression management and the construction of legitimacy. *Accounting, Auditing & Accountability Journal, 18*(3), 313-345. https://doi.org/10.1108/09513570510600729

Patten, D. M. (1991). Exposure, legitimacy, and social disclosure. *Journal of Accounting and Public Policy, 10*(4), 297-308. https://doi.org/10.1016/0278-4254(91)90003-3

Patten, D. M. (2002). The relation between environmental performance and environmenta disclosure: a research note. *Accounting, Organizations and Society, 27*(8), 763-773. https://doi.org/10.1016/S0361-3682(02)00028-4

Patten, D. M. (2019). Seeking legitimacy. *Sustainability Accounting, Management and Policy Journal, 11*(6), 1009-1021. https://doi.org/10.1108/sampj-12-2018-0332

Pencle, N., & Malaescu, I. (2016). *Corporate Social Responsibility Dictionary For Text Analytics*. https://provalisresearch.com/products/content-analysis-software/wordstat-dictionary/corporate-social-responsibility-text-analytics/

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*.

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: the case of college admissions essays. *PLoS One, 9*(12), e115844. https://doi.org/10.1371/journal.pone.0115844

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzalez, A., & Booth, R. J. (2007). *The Development And Psychometric Properties of LIWC2007* (1). The University of Texas at Austin.

Prado-Lorenzo, J. M., Dwyer, R., Rodríguez-Domínguez, L., Gallego-Álvarez, I., & García-Sánchez, I. M. (2009). Factors influencing the disclosure of greenhouse gas emissions in companies world-wide. *Management Decision, 47*(7), 1133-1157. https://doi.org/10.1108/00251740910978340

Reverte, C. (2009). Determinants of Corporate Social Responsibility Disclosure Ratings by Spanish Listed Firms. *Journal of Business Ethics, 88*(2), 351-366. https://doi.org/10.1007/s10551-008-9968-9

Riera, M., & Iborra, M. (2017). Corporate social irresponsibility: review and conceptual boundaries. *European Journal of Management and Business Economics, 26*(2), 146-162. https://doi.org/10.1108/ejmbe-07-2017-009

Rogers, J. L., Van Buskirk, A., & Zechman, S. L. C. (2011). Disclosure Tone and Shareholder Litigation. *The Accounting Review, 86*(6), 2155-2183. https://doi.org/10.2308/accr-10137

Roszkowska-Menkes, M., Aluchna, M., & Kaminski, B. (2024). True transparency or mere decoupling? The study of selective discloure in sustainability reporting. *Critical Perspectives on Accounting, 98,* 1-24. https://doi.org/10.1016/j.cpa.2023.102700

Ruiz-Blanco, S., Romero, S., & Fernandez-Feijoo, B. (2021). Green, blue or black, but washing–What company characteristics determine greenwashing? *Environment, Development and Sustainability, 24*(3), 4024-4045. https://doi.org/10.1007/s10668-021-01602-x

Sandberg, M., & Holmlund, M. (2015). Impression management tactics in sustainability reporting. *Social Responsibility Journal, 11*(4), 677-689. https://doi.org/10.1108/srj-12-2013-0152

Shakil, M. H. (2021). Environmental, social and governance performance and financial risk: Moderating role of ESG controversies and board gender diversity. *Resources Policy, 72*. https://doi.org/10.1016/j.resourpol.2021.102144

Stacchezzini, R., Melloni, G., & Lai, A. (2016). Sustainability management and reporting: the role of integrated reporting for communicating corporate sustainability management. *Journal of Cleaner Production, 136*, 102-110. https://doi.org/10.1016/j.jclepro.2016.01.109

Strike, V. M., Gao, J., & Bansal, P. (2006). Being good while being bad: social responsibility and the international diversification of US firms. *Journal of International Business Studies, 37*(6), 850-862. https://doi.org/10.1057/palgrave.jibs.8400226

Suchman, M. C. (1995). MANAGING LEGITIMACY: STRATEGIC AND INSTITUTIONAL APPROACHES. *Academy o/ Management Review, 20*(3), 571-610. https://doi.org/10.5465/amr.1995.9508080331
Talbot, D., & Boiral, O. (2018). GHG Reporting and Impression Management: An Assessment of Sustainability Reports from the Energy Sector. *Journal of Business Ethics, 147*(2), 367-383. https://doi.org/10.1007/s10551-015-2979-4
Tan, H.-T., Ying Wang, E., & Zhou, B. O. (2014). When the Use of Positive Language Backfires: The Joint Effect of Tone, Readability, and Investor Sophistication on Earnings Judgments. *Journal of Accounting Research, 52*(1), 273-302. https://doi.org/10.1111/1475-679x.12039
Tata, J., & Prasad, S. (2015). National cultural values, sustainability beliefs, and organizational initiatives. *Cross Cultural Management, 22*(2), 278-296. https://doi.org/10.1108/ccm-03-2014-0028
Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology, 29*(1), 24-54. https://doi.org/10.1177/0261927x09351676
Tay, D. (2021). Modelability across time as a signature of identity construction on YouTube. *Journal of Pragmatics, 182*, 1-15. https://doi.org/10.1016/j.pragma.2021.06.004
Tedeschi, J. T., & Melburg, V. (1984). Impression management and influence in the organization. In S. B. Bacharach, & E. J. Lawler (Eds.), *Research in the sociology of organizations* (pp. 31-58). JAI Press.
Van Caneghem, T., & Aerts, W. (2011). Intra-industry conformity in dividend policy. *Managerial Finance, 37*(6), 492-516. https://doi.org/10.1108/03074351111134718
Vasilescu, C., & Wisniewski, T. P. (2020). *Rebuilding a shattered reputation: Discretionary accruals and donations following corporate controversies*.
Verrecchia, R. E. (1983). Discretionary Disclosure. *Journal of Accounting and Economics, 5*, 179-194. https://doi.org/10.1016/0165-4101(83)90011-3
Villiers, C. d., & van Staden, C. J. (2011). Where firms choose to disclose voluntary environmental information. *Journal of Accounting and Public Policy, 30*(6), 504-525. https://doi.org/10.1016/j.jaccpubpol.2011.03.005
Weick, K. E. (1995). *Sensemaking in organizations*. Sage Publications.
Yan, B., Aerts, W., & Thewissen, J. (2019). The informativeness of impression management − financial analysts and rhetorical style of CEO letters. *Pacific Accounting Review, 31*(3), 462-496. https://doi.org/10.1108/par-09-2017-0063
Zhang, S., & Aerts, W. (2015). Management's performance justification and failure to meet earnings thresholds. *Accounting and Business Research, 45*(6-7), 841-868. https://doi.org/10.1080/00014788.2015.1048771

Appendix 1

To examine the discourse in CEO letters in SR we use an unsupervised machine learning method, i.e., topic modeling by latent Dirichlet allocation (LDA). To prepare our data for the analysis we convert the PDF files into plain text files using Python; tokenize text files (splitting sentences into words); stem terms (annotating every token for the base form) and remove stop-words using natural language processing (NLP) tools on Python. Then, we investigate topics in each CEO letters.

Topic modeling approach allows to scan a series of documents and find topic-based patterns within them (Fiandrino & Tonelli, 2021). The LDA method discovers new features in textual data, characterized by the probability of co-occurring words within documents, in which the topics are inductively labeled by researchers based on the words identified with (Blei et al., 2003; Brown et al., 2020). It does not require any pre-determined dictionaries or word-lists and relies on the frequent words appearing together that tend to be semantically related (Brown et al., 2020). Words are investigated based on their relative importance (weight) for the respected topic. While the same word may be present in different topics, its weight differs between topics. With this approach LDA accounts for polysemy and multiple meanings of a word depending on the context (Fiandrino & Tonelli, 2021). To avoid tokens that appear too frequently we ignore words that are more than in 90% of the documents and that are less than in 10% of the documents. In LDA, the optimal number of topics is unknown, and set by researchers manually, which can lead to different interpretations of the topics discovered (Huang et al., 2018). We identify the number of interpretable topics by measuring the perplexity of topic modeling, which assesses an LDA model's ability to predict word choices and used as a determinant of the number of topics (Huang et al., 2018). The lower perplexity indicates that the model is a better fit for the observed data and that the model gains less from increasing the number of topics (Dyer et al., 2017). We test and plot the perplexity scores for different number of topics, ranging from 1 to 100. We observe that perplexity scores are at the lowest for 7 topics. Table A-1 presents the topic-word probability matrix, according to which we label topics inductively.

The LDA assigns analyzed documents with scores that quantify the extent (between 0 and 1) to which each topic is discussed in each document based on the topic-word probability matrix, which is referred to as “topic loadings” or document-by-topic matrix. Using this data, we measure the intensity of sustainability discourse in each SR CEO letter.

Table A-1. Topics in CEO letters in SR

Topics	Top 10 words according to weight in	Citations from the most representative letters
Energy	energi, custom, power, company, technolog, electr, employe, commun, gener, year	*Given the public's growing desire for renewable and distributed energy resources beyond the traditional forms of electric, gas, and steam, we are preparing for a future that includes newer resources—including solar, wind, combined heat and power, fuel cells, and battery storage.*
Environment	sustain, product, water, reduc, energi, emiss, wast, use, environment, improv	*In 2013, [we] made significant progress toward our goal of achieving, by 2020, 15 percent reductions on a per-unit-of-production basis in energy, emissions, water and waste. In addition, we replaced our environmental mission and principles statement with a formal environmental policy…*
Community	commun, company, respons, people, world, busi, commit, make, work, year	*We make decisions and support causes that matter to the families we serve. Through our philanthropic platform (...) we have donated more than $51 million in 2016 to thousands of nonprofit organizations and causes across the country – organizations committed to making our communities stronger.*
Employees	oper, respons, busi, safety, continu, commun, perform, develop, stakehold, employe	*During 2015, we will continue to prioritise the health, safety, security and wellbeing of people while continuing to promote safe behaviours of our contractors.*
Financials	year, market, growth, new, custom, busi, continu, oper, wa, servic	*All these data make [us] the most profitable Spanish audiovisual company and one of the leading European media companies, thanks once again to the strength of our business model and to our excellent human team*

Sustainability-Commitment	sustain, group, company, develop, ha, year, social, energi, commit, global	*As a company that operates internationally, LEONI is committed to sustainable and responsible action. We have made it our mission to ensure that the company success is compatible with social and ecological principles.*
BusinessCase	sustain, busi, work, custom, product, year, make, develop, new, way	*But much more than that, in order to remain a successful business, we need to keep growing – and at the same time respect the planetary boundaries. So, there is no question that it makes clear business sense to invest in our sustainability.*

The identified topics can be divided into two main discourses, with (1) Energy, Environment, Community and Employees referring to sustainability-specific issues, i.e., sustainability-related content intensity; and (2) Financials, SustainabilityCommitment and BusinessCase focusing on general company's commitments, its approach to sustainability and how it relates to business success, i.e., business strategic case-related content intensity. This approach allows us to control for sustainability-related content intensity and business strategic-case-related content intensity. As the discourses represent probability and their sum equals one, there is a perfect inverse correlation (-1.0) between them. Thus, we use only one of the variables (strategic business case-related content intensity), namely StratDis, as a control in our regression models.

To validate our interpretation of the topics we conducted additional text analysis in LIWC using CSR dictionary proposed by Pencle and Malaescu (2016). The authors generated (mutually non-exclusive) wordlists related to four CSR dimensions: (1) employee; (2) environment; (3) human rights; and (4) social and community. Examples of words divided into four categories are presented in Table A-2.

Table A-2. CSR word scores: dimensions and examples of words

Dimension	Examples of words/expressions
Human Rights (CSR_humanrights)	Disadvantage, equality, ethical, ethnic_diversity, gender, honesty, human_development, inclusive, nationality, races
Employee (CSR_employee)	Employee, employee_involvment, employee_safety, employee_welfare, employer, goal, health, jobs, mortality, wage
Environment (CSR_environment)	Conserve, energy_efficiency, environmental, epa, facility, global_warming, green_building, hazardous_waste, hybrid, organic
Social and Community (CSR_socialcommunity)	Community, CSR, government, human_being, impact_on_community, indigenous, involve, not_for_profit, open, orphan

Source: (Pencle and Malaescu, 2016)

We scanned our sample reports for the words related to different CSR dimensions and obtained scores (*CSR_i,* where *i* represents the dimension) representing percentage of words from a given list in the total number of words in the document. Next, we run a pairwise correlation between CSR word scores and topic scores generated through the LDA. The results are presented in Table A-3.

Table A-3. Pairwise correlations between LDA topic scores and CSR word scores

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) Energy	1.000										
(2) Environment	-0.067*	1.000									
(3) Financials	-0.108*	-0.134*	1.000								
(4) Community	-0.213*	-0.088*	-0.184*	1.000							
(5) Sustain. commit.	-0.171*	-0.196*	-0.030	-0.342*	1.000						
(6) Business case	-0.237*	-0.125*	-0.139*	-0.241*	-0.022	1.000					
(7) Employees	-0.173*	-0.126*	-0.180*	-0.297*	-0.163*	-0.143*	1.000				
(8) CSR_humanrigh.	-0.052	-0.123*	-0.277*	0.293*	-0.189*	-0.305*	0.414*	1.000			
(9) CSR_employee	-0.114*	-0.119*	-0.285*	0.278*	-0.159*	-0.251*	0.419*	0.794*	1.000		
(10) CSR_socialc.	-0.043	0.087*	-0.443*	0.241*	-0.133*	-0.112*	0.233*	0.620*	0.659*	1.000	
(11) CSR_envir.	0.126*	0.429*	-0.417*	-0.071*	-0.049	-0.225*	0.225*	0.254*	0.357*	0.526*	1.000

*Correlations significant at p<0.01

As shown in Table A-3, we find positive correlation between CSR_environment and both Environment and Energy, with less profound relationship in the case of the latter. Our community topic (Community) seems to overlap with three socially-oriented scores for human rights, employee, and social and community. Also, our Employee topic is correlated with these dimension, mostly however with human rights and employee, what is in line with our interpretation of this topic. Financials representing financial-accounting content with no reference to sustainability is, not surprisingly, negatively correlated with all CSR word scores. We find negative correlation with all four CSR dimensions also for Sustainability Commitment Business Case. This suggests that while these topics, in contrast to Financials, include discussion on sustainability as such, they do not refer to any specific aspect of sustainable business. These results support the distinction drawn between sustainability-specific topics and those focused on strategic business discourse.

References List (Appendix)

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research, 3*, 993-1022.

Brown, N. C., Crowley, R. M., & Elliott, W. B. (2020). What Are You Saying? Using topic to Detect Financial Misreporting. *Journal of Accounting Research, 58*(1), 237-291. https://doi.org/10.1111/1475-679x.12294

Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics, 64*(2-3), 221-245. https://doi.org/10.1016/j.jacceco.2017.07.002

Fiandrino, S., & Tonelli, A. (2021). A Text-Mining Analysis on the Review of the Non-Financial Reporting Directive: Bringing Value Creation for Stakeholders into Accounting. *Sustainability, 13*(2). https://doi.org/10.3390/su13020763

Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R. (2018). Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach. *Management Science, 64*(6), 2833-2855. https://doi.org/10.1287/mnsc.2017.2751

Pencle, N., & Malaescu, I. (2016). *Corporate Social Responsibility Dictionary For Text Analytics*. https://provalisresearch.com/products/content-analysis-software/wordstat-dictionary/corporate-social-responsibility-text-analytics/

CHAPTER IV

STUDY 3: FEASIBILITY ANALYSIS OF MACHINE LEARNING FOR PERFORMANCE-RELATED ATTRIBUTIONAL STATEMENTS

Abstract

We investigate the feasibility of machine learning methods for attributional content and framing analysis in corporate reporting. We test the performance of five widely-used supervised machine learning classifiers (naïve Bayes, logistic regression, support vector machines, random forests, decision trees) in a top-down three-level hierarchical setting to (1) identify performance-related statements; (2) detect attributions in these; and (3) classify the content of the attributional statements. The training set comprises manually coded statements from a corpus of management commentary reports of listed companies. The attributions include both intra- and inter-sentential attributional statements. The results show that for both intra- and inter-sentential attributions, F1-scores of our most accurate classifier (i.e., support vector machines) vary in the range of 76% up to 94%, depending on the identification, detection and classification levels and the content characteristics of attributions. Additionally, we assess the hierarchical performance of classifiers, providing insights into a more holistic classification process for attributional statements. Overall, our results show how machine learning methods may facilitate narrative disclosure analysis by providing a more efficient way to detect and classify performance-related attributional statements. Our findings contribute to the accounting and management literature by providing a basis for implementing machine learning methodologies for research investigating attributional behavior and related IM.

1. Introduction

Narrative disclosures have been the subject of many studies in the accounting and management literature. The narrative sections in corporate reports tend to complement the financial statements with discretionary information and are considered a significant means of communication (Henry & Leone, 2016; Li, 2010a; Shirata & Sakagami, 2008; Yang et al., 2018). In the narrative sections, companies often provide management's view on the context within which to interpret the performance, position, and progress of the company (Aerts & Tarca, 2010; Lewis & Young, 2019; Li, 2010a; Shirata et al., 2011). Examples of such narratives are the Directors' report, the Management Discussion and Analysis (MD&A – USA), the Operating and Financial Review (OFR – UK), the CEO letter to shareholders. As the content of these narratives is largely discretionary, they are believed to serve informational (signaling), as well as self-presentational purposes (Merkl-Davies et al., 2011).

Research in accounting narratives examines both informational and presentational dimensions (Aerts, 2005; Aerts & Yan, 2017). Presentational studies are related to IM and primarily focus on selectivity and bias in the content patterns of the narrative disclosures. Such presentational tendencies may manifest themselves relative to the choice of thematic content, how benchmarking occurs, the tone of the language used, the readability of the text, the use of emphasis in presentation, and how performance and events are explained. Explanation occurs when management moves from just providing information on performance to matters of meaning, relationships, causes, and reasons. It is here that attributional statements come in. An attributional statement can be defined as "*a phrase or a sentence in which a corporate event or performance outcome is linked with a reason or a cause for the event or outcome*" (Aerts, 2005). Attributional disclosure studies demonstrated a robust tendency to attribute positive events or outcomes to the company's own actions or resources and negative outcomes to external events or chance factors (such as economic climate, inflation, market prices, government policy, weather) (Aerts, 2005; Clatworthy & Jones, 2003; Hooghiemstra, 2003). This explanation pattern is considered self-serving as it tends to define situations to the company's own benefit. With this self-presentational bias in mind, attributional disclosure studies have often adopted an IM focus. In this research, different types of attributional structures have been identified, such as excuses, justifications, causality denials, enhancements, entitlements and tautological (accounting) explanations.

Textual analysis of the narratives has been a core issue in narrative disclosure studies. The linguistic features studied and the techniques used to extract these features differ, however. The use of automated text analyses programs, such as *General Inquirer* (GI), *Diction,* or *Linguistic Inquiry and Word Count* (LIWC), has become popular over the last two decades (Li, 2010a). These software packages are partly grounded in computational linguistics and allow the use of pre-defined domain-specific word lists or psychosocial dictionaries, such as words related to positive and negative emotion, words related to personal pronouns or to cognitive processes, to measure one-dimensional text features such as topical focus, tone, self-referencing or the presence of argument (Tausczik & Pennebaker, 2009). However, when the textual content and features of interest are more complex and multidimensional, automatic information retrieval becomes more difficult (El-Haj et al., 2019a). Manual content analysis is still favored when more subtle difference in or shade of content or expression is important (Li, 2010a). With manual content analysis, human coders score and categorize selected textual data independently based on pre-defined rules and coding schemes (Chakraborty et al., 2014; El-Haj et al., 2019a; Săndulescu, 2019). Automated information retrieval becomes more challenging when the narrative features of interest are more phrase- or sentence-based than word-based, which is the case for management's attributional statements (Aerts, 2005). Moreover, various linguistic devices can be used to indicate attribution or causality, such as conjunctions or verbs. Attributional statements can be confined to one sentence ("*intra-*

sentential") but may also span different (neighboring) sentences ("*inter-sentential*") (Kruengkrai et al., 2017; Oh et al., 2013; Zhao et al., 2016). This diversity in attributional format explains why researchers usually adopt a manual content analysis approach to identify and qualify attributions (e.g., Baginski et al., 2000; Clatworthy & Jones, 2003; Jayamohan et al., 2017; Kimbrough & Wang, 2014; Ogden & Clarke, 2005; Săndulescu, 2019). As manual content analysis is a time-consuming and resource-intensive process, the volume of available and relevant textual data in narrative reports usually exceeds available human capacity and resources to process disclosures manually (Chakraborty et al., 2014; El-Haj et al., 2016; Fisher et al., 2010). Therefore, studies that adopt manual text analysis tend to have a smaller sample size (Fisher et al., 2010; Li, 2010a). Furthermore, manual text analysis may involve individual views and interpretations of coders in the coding process and increase subjectivity of data collection (Aerts & Tarca, 2010; Chakraborty et al., 2014; El-Haj et al., 2016). These factors may limit the validity, representativeness and reliability of the analysis and leverage the potential of machine learning algorithms as an efficient means for text classification of larger datasets (Fisher et al., 2010).

This study investigates the feasibility of applying machine learning methods for attributional content and framing analysis in accounting narratives. Although machine learning techniques have been drawing increasingly more attention as a tool for textual analysis, to date only a limited number of studies in accounting and management have used machine learning algorithms to classify facets of attributional statements in narrative disclosures (e.g., El-Haj et al., 2016; Li, 2010b; Li, 2010c; Walker et al., 2020), but these leave many questions on the feasibility of using classifiers for an automated coding process for the full repertoire of attributional statements. First, the majority of previous studies have employed machine learning techniques as a supplementary methodology for classification of linguistic features after attributional statements were selected based on a dictionary approach (e.g., Li, 2010c;

Walker et al., 2020). Nevertheless, to use machine learning algorithms for a full scope narrative coding process, securing the accuracy of initial identification steps is required, as, if neglected, errors at higher identification levels are carried over to the lower levels (Kiritchenko et al., 2005). We assess the performance of five widely-used machine learning algorithms in a top-down three-level hierarchical structure to (1) identify performance-related statements; (2) detect attributions; and (3) classify statements according to attributional characteristics. By testing the performance of the top-down three-level hierarchical classification, we apply a more holistic approach to assess the general performance of classifiers for the entire coding process and identify the tasks that create most difficulty for the automatization of the classification process. Second, prior studies, that adopted machine learning algorithms for attributions, have mainly focused on sentence-level classification, where inter-sentential attributions are neglected (Lamm et al., 2018). Neglecting inter-sentential attributions is an important caveat, as Aerts and Tarca (2010), for example, show that half of the attributions in their sample are inter-sentential. We, therefore, compare the performance of five widely-used machine learning algorithms at both intra- and inter-sentential levels and show the best-fit classifiers for automated classification of performance-related attributional statements task. Lastly, up to now, prior research that used machine learning algorithms has focused mainly on the basic self-serving attribution bias (entitlements versus excuses) using two key features of attributions: valence of the explained effect and locus of causality. This focus ignores other fine-grained attributional structures which are common in narrative disclosures, such as enhancements, causality denials, justifications, and formal language explanations, and which have been shown to be used for assertive and defensive IM. We extend the literature by analyzing the performance of classifiers on four distinct characteristics of attributed effects (i.e., nature of effect, valence of effect, quantitative/qualitative nature of effect, level of effect), and four content features for causal factors (i.e., direction of cause-effect relationship, qualitative/quantitative nature of cause, nature of cause, locus of causality). Taken together, these eight features of attributional statements allow to identify and measure attributional structures beyond entitlements and excuses.

To train machine learning algorithms and perform our tests, we use a corpus of management commentary reports (MD&A or operating and financial review (OFR) reports) of listed companies from four common law countries (USA, Canada, UK, and Australia) and five industries (building materials, food processors, pharmaceuticals, biotechnology, and retail). The corpus consists of approximately 1.21 million words and comprises a total of 4,585 annotated attributional statements across 172 companies. Our results demonstrate that F1-scores for identification, detection and classification levels vary in the range of 58% up to 94%, depending on the specific classifier being used and the content and style-related characteristics of attributions. Additionally, we test the three-level hierarchical performance on locus of causality, i.e., external versus internal attributions, using a gold standard dataset (i.e., human-annotated, external data) by El Haj et al. (2016), which results in an F1-score of 61%. In supplementary analyses, we perform external validity checks for hierarchical classification, which gives an error rate of 20%, on average, depending on different content and style-related textual characteristics that we investigate in the third level. We also compare the performance of our best-performing machine learning algorithm with automated text analyses programs (LIWC and the Coh-Metrix) on causality detection (level 2), which is the most difficult part in the hierarchical setting. The comparison provides evidence that machine learning methods may outperform the traditional automated text analyses programs for detecting attributions.

Overall, our paper provides additional insights into an alternative methodology for content and style analysis of attributional statements in narrative disclosures. These insights may be of assistance to researchers who intend to investigate fine-grained attributional behavior and related IM in larger samples of corporate disclosures. The next sections of the paper are organized as follows: Section 2 discusses narrative disclosures and IM; as well as methodological approaches that have been used in the literature. Section 3 introduces methodology and data. Section 4 provides the results. Finally, section 5 presents a discussion of the findings and our conclusions.

2. Literature Review

This section provides background information on the relevance of attributional statements for corporate narrative disclosure research and on the textual analysis methods used to investigate attributions in corporate narratives. Next, we briefly review how machine learning methods have been used in prior research on corporate narratives.

2.1. Management commentary and narrative disclosure research

Management commentary reports such as a CEO letter to shareholders, an MD&A report or a directors' report, are an integral component of a company's periodic financial communication repertoire. They accompany the traditional financial statements with a narrative description of the company's accomplishments and performance outcomes in the period under review, of significant events that affected the company's current financial condition and may include prospective statements regarding future developments (Aerts & Tarca, 2010; IASB, 2010). They often elaborate a framing context, whereby events and performance are explained and put into context (El-Haj et al., 2019b; El-Haj et al., 2016; Lewis & Young, 2019; Li, 2010a; Shirata et al., 2011). Prior research shows that narratives in corporate disclosure can significantly affect market reactions and investment decisions (Baginski et al., 2004; Davis et al., 2012; Huang et al., 2014; Merkl-Davies & Brennan, 2007). Because narrative disclosures are to a large extent discretionary, they are deemed to be instrumental in controlling and directing the perception of the reader and, thus, relevant for corporate IM (Aerts, 2005; Merkl-Davies et al., 2011). Merkl-Davies and Brennan (2007) categorize verbal IM behaviors as concealment and attributional strategies.

2.1.1. Concealment

Concealment can be achieved through thematic manipulation, e.g., the use of tone to emphasize good news, and through syntactical manipulation, e.g., the use of complicated language to obfuscate bad news (Melloni et al., 2016). A large body of prior narrative reporting research has focused on tone manipulation (e.g., Abrahamson & Park, 1994; Davis & Tama-Sweet, 2012; Du & Yu, 2020; Henry, 2006, 2008; Li, 2010b; Melloni et al., 2016) and readability (e.g., Lawrence, 2013; Lehavy et al., 2011; Li, 2008; Loughran & McDonald, 2016; Rennekamp, 2012).

Research in this area usually relies on computer-aided methods. Software to apply readability formulae, measuring readability in terms of word and sentence length, is easily accessible. Tone analysis (and thematic content inquiry in general) typically uses an automated dictionary approach. A dictionary is a tabulated collection of keywords or phrases with an associated feature (Guo et al., 2016; Henry, 2006; Loughran & McDonald, 2011). Words (including word stems) from pre-defined word lists are counted and reported as a percentage of total words using software, such as *Harvard's General Inquirer* (GI), *Diction*, or *Linguistic Inquiry and Word Count* (LIWC) (Boyd et al., 2022; Henry, 2006). In the literature, psychosocial dictionaries have been used to measure tone, but also cognitive complexity, self-referencing, and causal words (e.g., Davis & Tama-Sweet, 2012; Huang et al., 2014; Im et al., 2013; Merkl-Davies et al., 2011). Dictionary-based text analysis improves the capability of working with larger sample sizes (Li, 2010a) and it allows researchers to create custom dictionaries according to their focus of interest (Pennebaker et al., 2015). Indeed, standard dictionaries are not always accurate for specific language domains (Loughran & McDonald, 2011). For example, language in financial disclosures is often jargon-like and the meaning of the words used in those disclosures may well differ from everyday language, e.g., the specific meaning of the word 'liability' in financial disclosure. In this regard, Loughran and McDonald (2011) created their own custom word-lists to measure tone, modal words and uncertainty in financial disclosures, as they demonstrated that standard dictionaries, such as those applied in *Harvard-IV-4*, tend to misclassify common words in financial texts because of their domain-specific meaning in accounting language. Similarly, Henry (2008) created domain-specific word-lists to analyze tone and topical content in earnings press releases.

Because of the advantages of generalizability and customization, numerous studies have conducted dictionary-based text analysis (e.g., Boritz et al., 2013; Davis & Tama-Sweet, 2012; Huang et al., 2014; Im et al., 2013; Merkl-Davies et al., 2011; Zhang & Aerts, 2015; Zhang et al., 2019). The dictionary approach has, however, its limitations (El-Haj et al., 2019a; Li, 2010a). As previously stated, customization of domain-specific language may be challenging (Li, 2010a), but its main disadvantage lies in the use of words as a unit of measurement. It tends to neglect the context of words and to ignore touches of slang, irony, sarcasm, and idioms (Guo et al., 2016; Henry, 2008; Li, 2010a; Loughran & McDonald, 2016; Tausczik & Pennebaker, 2009). When the term "*increase"* is used with the word *"sales",* the tone is generally seen as positive. However, when it is used in a phrase as "*increase cost"*, the tone of the phrase is more likely to be perceived as negative. Another bottleneck for a word-based approach to measure tone is 'double negatives'. Words, such as "*not*" and "*bad*" are negative when used alone, yet the tone changes to positive when they are used together "*not bad*". Additionally, polysemous words (i.e., terms with multiple meanings) or parts of company names, that are accounted for in a dictionary category, such as "*Best Buy Co.*" may cause further lexical ambiguity. By focusing on words, a dictionary approach tends to ignore meaning embedded in phrases and sentences.

2.1.2. Attribution

Another stream of verbal IM identified by Merkl-Davies and Brennan (2007) relates to attribution-based communication. Applied attribution theory studies have been popular in management commentary research, especially because narrative elaboration of the 'why' and 'how' of reported accounting data is where new information, supplementary to the financial statement data, is created (e.g., Jayamohan et al., 2017; Săndulescu, 2019; Walker et al., 2020). Most performance explanations in management commentary can be considered attributions as they elaborate on a relationship between an antecedent and a performance-related consequence (Aerts, 1994, 2005; Aerts & Cheng, 2011). Attribution, more than concealment, is based on the meaning of the information disclosed in a string of words, which makes it more challenging to examine (Cho et al., 2010; El-Haj et al., 2016). The information retrieval process of attributions requires detailed investigation of causal inferences ('causality mining') and is not easily automatized (Aerts, 2005; Zhao et al., 2016). Attributions can use a range of linguistic devices to construct causal inferences, such as causal connectors ("*thus*", "*therefore*"), causal connecting phrases ("*because of*", "*as a result of*")[19], or verbs that refer to causality, such as "*affect*" or "*force*" (Aerts, 2005; Walker et al., 2020). Depending on the causality connectives, attributions can be constructed both as "*intra-sentential*" (within a sentence) and "*inter-sentential*" (between neighboring sentences) (Oh et al., 2013; Zhao et al., 2016). Example statement 1 and example statement 2 below illustrate intra- and inter-sentential attributions:

> Example statement 1: "*Total net sales increased 6% or $14.3 billion during 2020 compared to 2019, primarily driven by higher net sales of Services and Wearables, Home and Accessories"* (Apple, 2020 Annual Report, p. 20).
>
> Example statement 2: "*The Women's Health franchise sales were $0.9 billion in 2020, a decrease of 8.6% as compared to the prior year. The decline was primarily driven by COVID-19 impacts"* (Johnson & Johnson, 2020 Annual Report, p. 22).

[19] Garzone (2006, pp. 91-92) classifies causal connectors into three: (1) subordinating (because/since/as); (2) prepositional (due to/as a result of/because of); (3) adverbial (as a result/therefore/consequently). As she argues, the choice of causal connectors may be related to the different rhetorical effects of causal connectors, i.e., semantically strong versus weak causality (Leibbrand, 2015).

In *example statement 1*, attribution is constructed in one sentence, but in *example statement 2*, the performance outcome, i.e., "*sales decrease"*, is part of the first sentence, and the reason of the performance outcome, i.e., "*COVID-19 impacts*" is added as the second sentence.

Prior research investigating event causality tends to conduct sentence-level analysis (e.g., El-Haj et al., 2016). Considering inter-sentential attributions is, however, essential when investigating attributions. The manually annotated dataset by Aerts and Tarca (2010), for example, includes 2,545 intra-sentential attributions and 2,040 inter-sentential attributions. This indicates that focusing only on intra-sentential attributions would leave half of the attributions uninvestigated. Walker et al. (2020) is a notable exception investigating the inter-sentential properties of attributional statements using machine learning algorithms. They conduct a study at a tri-sentence level, i.e., three sentences consisting of a performance-related sentence and two immediately adjacent sentences.

Explaining performance outcomes with attributions can be used opportunistically to affect reader perception (Baginski et al., 2011; Melloni et al., 2016; Merkl-Davies & Brennan, 2007). While some studies emphasize the motives behind attributional IM, such as overconfidence (Li, 2010c; Libby & Rennekamp, 2012), or public scrutiny (Aerts, 2005), the vast majority focus on attributional features and how they are instrumentalized (e.g., Baginski et al., 2000; Clatworthy & Jones, 2003; Jayamohan et al., 2017; Kimbrough & Wang, 2014; Li, 2010c; Libby & Rennekamp, 2012; Săndulescu, 2019). Attributional disclosure studies demonstrated that attributions can serve both assertive, e.g., enhancing the effect of positive outcomes, and defensive, e.g., downplaying negative outcomes, purposes (Aerts & Cheng, 2011), and reported a robust tendency to take credit for positive events or performance outcomes, while blaming outside factors for negative outcomes. For example, Baginski et al. (2000) find the existence of self-serving attributional bias in earnings forecasts, where management uses more external (internal) reasons for bad (good) forecast news. Kimbrough and Wang (2014) find that firms suffer less market penalties after providing defensive attributions and earn greater market rewards with assertive attributions. Moreover, Aerts (1994, 2005) and Aerts and Tarca (2010) identify managers' tendency to explain a firm's negative performance using more technical terms or accounting logic arguments, while positive firm performance tends to be explained in more direct causal language. They further decompose attributional behavior in specific assertive and defensive formats, such as entitlements, enhancements, excuses, causality denials, justifications, and formal language explanations, which reveal deeper insights into the differentiated use of attributional statements than focusing only on locus of causality[20].

Unlike studies investigating concealment in corporate disclosures, empirical studies that focus on attributions mostly adopt a manual text analysis approach due to the multicomponent (antecedent/consequence) identity of attributions, their phrase-based nature and the multiple linguistic devices that may be used to express attributional patterns (Aerts, 1994, 2005). These features make it more difficult to automatically process attributional information (El-Haj et al., 2019a). Manual text analysis is considered a precise and well-tailored method (Li, 2010a), that requires multiple human coders to score text files independently to categorize the selected textual data based on pre-defined rules (Chakraborty et al., 2014; El-Haj et al., 2019a). It necessitates transparency on training examples, coding rules and instructions, and procedures to establish inter-coder reliability (Aureli, 2017; Krippendorff, 2013). To date, the majority of previous studies have investigated attributional IM through manual text analysis (e.g., Aerts, 1994, 2005; Aerts & Cheng, 2011; Aerts & Tarca, 2010; Baginski et al., 2004; Brühl & Kury, 2016; Clatworthy & Jones, 2003; Jayamohan et al.,

[20] Assertive attributional IM includes entitlements, i.e., the attribution of positive outcomes to internal causes, and enhancements, i.e., highlighting positive outcomes in spite of negative external circumstances. Defensive attributional IM includes excuses, i.e., the attribution of negative outcomes to negative external factors, justifications, i.e., accepting the responsibility of a negative outcome but using this as a step to achieve higher goals, and causality denials, i.e., denying the responsibility of a negative outcome (Aerts, 2005; Aerts & Cheng, 2011; Săndulescu, 2019; Scott & Lyman, 1968).

2017; Kimbrough & Wang, 2014; Melloni et al., 2016; Ogden & Clarke, 2005; Rosenkranz & Pollach, 2016; Săndulescu, 2019). For example, Aerts (1994; 2005) manually identified and coded attributional statements in management commentaries to explore patterns in performance explanations. Different coders identified and qualified the statements independently. Attributional statements were coded based on a number of characteristics, such as locus of causality, valence of effect and outcome, and nature of explanation. Clatworthy and Jones (2003) manually coded good and bad news, and whether the news was attributed internally (to management) or externally (to other factors). Aerts and Tarca (2010) manually coded characteristics of explanatory statements such as nature, valence, locus of causality and time orientation. These detailed characteristics of the statements were then used to identify and measure specific attributional structures (entitlements, excuses, causality denials, etc.).

The main drawback of manual text analysis is the tension between the sheer volume of available qualitative data in narrative reports and the limited human capacity to manually process them (El-Haj et al., 2019a), which raises concerns over time consumption and coding cost (Chakraborty et al., 2014; El-Haj et al., 2016; Fisher et al., 2010), sample size (Chakraborty et al., 2014; Fisher et al., 2010; Li, 2010a), replicability and generalizability (Li, 2010a). These factors may limit the representativeness and power of studies using manual text analysis, where computer-based analysis may offer an opportunity to overcome these limitations (Fisher et al., 2010; Li, 2010a). For example, Fisher et al. (2016) show a substantial increase in sample size when computer-aided methods are used.

Some studies have used a dictionary-based approach for automated attributional analysis (e.g., Koo et al., 2017; Zhang & Aerts, 2015; Zhang et al., 2019), yet these studies have been limited to applying causal word and causal connector dictionaries. They essentially measure causal language intensity in corporate narratives, but lack in elaborating the full range of attributional content and their specific framing patterns (El-Haj et al., 2019a).

2.2. Machine learning methods and their use in narrative disclosure research

Textual files are considered to be unstructured data, lacking a sort of structure such as rows and columns (Provost & Fawcett, 2013, p. 250; Weiss, 2005, p. v). A common problem that researchers face when analyzing narrative disclosures is data reduction, which is transforming a large amount of unstructured data into workable numerical values. To capture content and style-related features of text, scholars have been following different methodologies to generate replicable and valid inferences from texts (Krippendorff, 2013). Machine learning can be an alternative methodology for researchers to identify thematic content and writing style and may overcome the disadvantages of traditional content analysis (El-Haj et al., 2019a; Li, 2010a; Van Atteveldt et al., 2021). It involves a collection of methods for extracting (predictive) models from data, and it arose as a subfield of artificial intelligence (Fisher et al., 2016; Provost & Fawcett, 2013, p. 39). In the accounting and finance literature, there has been a growing interest in using machine learning algorithms to study narrative disclosures (Boritz et al., 2013; El-Haj et al., 2019a; Fisher et al., 2016).

To prepare textual files for predictive statistical models, several pre-processing steps are required (Bach et al., 2019; Bickel, 2017; Chen, Wu, Chen, Li, & Chen, 2017; Ignatow & Mihalcea, 2018, p. 52; Liew et al., 2014; Szekely & Vom Brocke, 2017). These text pre-processing steps consist of sentence segmentation, tokenizing (splitting sentences into words), lemmatizing or stemming (annotating every token for the base form), and removing stop words, such as articles and prepositions (Bach et al., 2019; Bickel, 2017; Ignatow & Mihalcea, 2018, p. 52). To analyze cleaned text data, natural language processing (NLP) models, such as "bag-of-words" are used (Bach et al., 2019)[21]. NLP models are crucial to split text into a vector of terms in a document-term-matrix, that represents the frequency of each token per document

[21] Bag-of-words is an NLP method used to parse documents into a matrix composed of words and word count vectors, i.e., term-document matrix based on word frequencies (Loughran & McDonald, 2016).

(Bickel, 2017). The vectors of term counts are required to be normalized, as the term count is strongly tied to document length (Loughran & McDonald, 2016). For example, TF-IDF (term frequency-inverse document frequency) is a common term weighting method used in text mining (Liu et al., 2017; Loughran & McDonald, 2011). After the text pre-processing steps, machine learning methods can be applied (Bickel, 2017). Machine learning methods are divided into two approaches: (1) unsupervised approach and (2) supervised approach.

Unsupervised learning methods rely on algorithms that learn how to group raw data automatically without human intervention (El-Haj et al., 2019a). They are used for tasks like clustering and topic modeling (Guo et al., 2016; Provost & Fawcett, 2013, p. 24). Topic models can be created with matrix factorization methods, such as Latent Dirichlet Allocation (LDA) (Blei et al., 2003; Guo et al., 2016; Provost & Fawcett, 2013, p. 265). LDA creates topics based on the probability of words co-occurring within documents (Blei et al., 2003; Dyer, Lang, & Stice-Lawrence, 2017). For example, Dyer et al. (2017) use LDA to examine managerial topics that caused longer financial disclosures over the last decade. As another example, Huang et al. (2018) adopt a topic modeling approach and investigate the thematic differences between narrative analyst reports and conference calls.

Most disclosure studies in the accounting and finance literature that apply machine learning algorithms, use supervised machine learning models to extract information from textual data at the sentence or document level. A supervised model requires a set of manually classified data for classification tasks (El-Haj et al., 2019a; Provost & Fawcett, 2013, p. 24; Sebastiani, 2002). Once the classes are constructed, the initial corpus is split into three sets: (1) a training set, (2) a test set, and (3) a validation set, which are used to tune the parameters of classifiers and to evaluate the effectiveness (Sebastiani, 2002). To evaluate the effectiveness of the models, the *K*-fold cross-validation method is commonly used, where both training and test sets are split into *K* equal-sized parts (El-Haj et al., 2019a). The effectiveness of a classifier is obtained by averaging the results of *K* different classifiers (Sebastiani, 2002). Naïve Bayes, logistic regression, support vector machines, random forests and decision trees are supervised learning algorithms frequently used in business studies (Provost & Fawcett, 2013, p. 43).

The naïve Bayes classifier is a relatively simple probabilistic classifier (El-Haj et al., 2019a; Sebastiani, 2002; Van den Bogaerd & Aerts, 2011). Li (2010c) applied the naïve Bayes approach in examining tone and content of forward-looking statements in MD&A. After training the model with 30,000 manually coded forward-looking statements, Li (2010c) classifies the tone and content of 13 million forward-looking statements. Huang et al. (2014) use the naïve Bayes machine learning approach to classify tone in analyst reports. Sprenger et al. (2014) train the naïve Bayes algorithm with 2,500 tweets as buy, hold, and sell signals for stock-related messages, and investigate the relationship between tweet sentiment and stock returns. Similarly, Neuenschwander et al. (2014) perform sentiment analysis using the naïve Bayes machine learning classifier on Brazilian stock market news and tweets. Although the naïve Bayesian model is a popular machine learning tool in disclosure studies, other supervised models like support vector machines, random forests, and logistic regression may well outperform the naïve Bayesian model (Antweiler & Frank, 2004; El-Haj et al., 2016; Goel et al., 2010). Researchers have been comparing results of applying different machine learning algorithms to determine the best-fit model for narrative disclosures as there is little guidance for reasoned choice in the literature (El-Haj et al., 2019a). El Haj et al. (2016) compare the performance of support vector machines, logistic regression, random forests, and naïve Bayes classifiers to classify performance and non-performance-related sentences in MD&A reports. The results reveal that the naïve Bayes classifier is the least accurate among the classifiers tested. Humpherys et al. (2011) and Goel et al. (2010) report conflicting results. Humpherys et al. (2011) analyze word quantity and word diversity of 10-K's, to train machine learning algorithms (support vector machines, decision trees, and naïve Bayesian models) and report that decision trees and naïve Bayes are the most accurate algorithms for distinguishing fraudulent from non-fraudulent 10-K's[22]. Goel et al. (2010), investigating linguistic features such as passive-active voice, readability, sentence, and word length to identify and classify fraudulent annual reports[23], show that support vector machines provide higher prediction accuracy than the naïve Bayes classifier.

Few researchers have attempted to apply machine learning algorithms to automatically classify attributions (El-Haj et al., 2016; Li, 2010c; Walker et al., 2020). Li (2010a) examines managers' self-referencing tendency in attributional sentences. He identifies and extracts attributional sentences with causative words based on the LIWC dictionary and uses a naïve Bayesian algorithm to classify tone (positive, negative, and neutral) of the attributional sentences. El-Haj et al. (2016) investigate the main self-serving attribution bias in UK preliminary earnings announcements using machine learning algorithms. Comparing the performance of four different supervised machine learning algorithms, the authors first classify performance and non-performance sentences with 70% accuracy, then classify performance statements based on sentence tone (positive, negative, neutral), attribution (internal, external), and attribution tone (positive, negative, neutral) with 79% accuracy, both of which are lower compared to manual inter-coder reliability. In a more recent study, Walker et al. (2020) test the performance of four different supervised machine learning algorithms for self-serving attributional bias, i.e., tone (positive, negative, neutral, unclear) and attribution type (internal, external) and compare their results with an approach using dictionaries. They suggest that automated text analysis tools have certain limitations, compared to manual content analysis. In our paper, we follow a similar approach to test the feasibility of machine learning algorithms, but we differ in testing the application of a hierarchical approach and in investigating a more

[22] Humpherys et al. (2011) identify fraudulent reports from the Securities and Exchange Commission's Accounting and Auditing Enforcement Releases.
[23] Goel et al. (2010) define fraudulent reporting as "*fraud that had affected 10-Ks through material manipulation, mispresentation, or failure to disclose material facts*".

diverse set of attributional features that allow us to measure attributional structures beyond the classic self-serving attributional bias.

The next part of this paper elaborates on the data we used to train our machine learning models and the methodological steps taken to arrive at an assessment of the feasibility of an automated attributional statement coding process.

3. Methodology and Data

As mentioned earlier, our study follows a three-level hierarchical classification process to simulate the attributional statement coding process (Figure 1). The first level contains two classes: (1) performance-related statements; and (2) non-performance-related statements. We train machine learning algorithms to extract (identify) performance-related statements. As not all performance-related statements are explanatory, at the second level, we detect performance-related statements with attributions by applying a second binary classification: (1) performance-related statements with attributions; and (2) performance-related statements without attributions. In this step, we extract attributional statements with an attributional statement to cover most attributional structures that occur in management commentary. Finally, in the third level, we train the machine learning algorithms to capture specific textual characteristics of the identified attributional statements, focusing on both content and style features (i.e., nature of effect, valence of effect, quantitative/qualitative nature of effect, level of effect, direction of cause-effect relationship, qualitative/quantitative expression of cause, nature of cause, locus of causality).

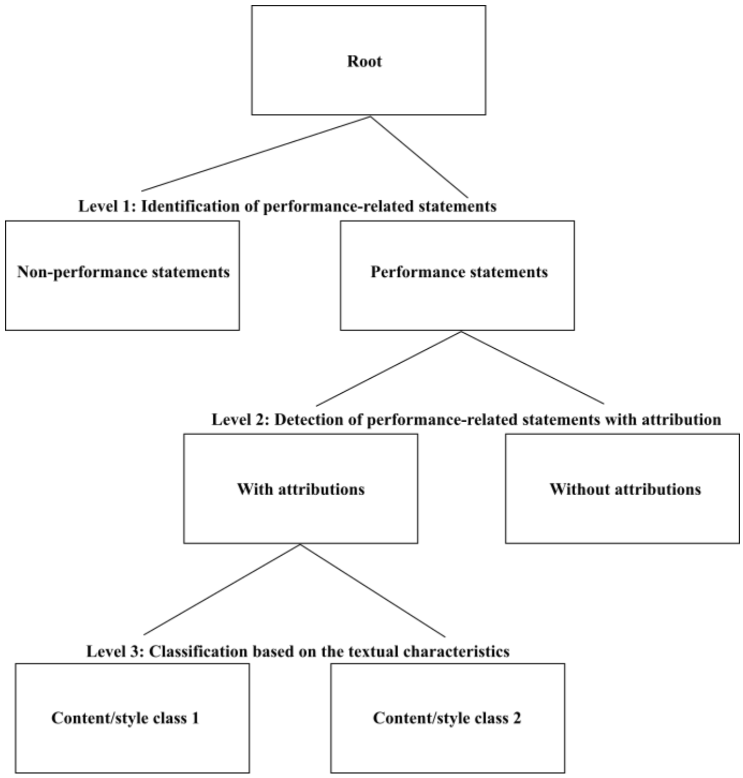

Figure 1: Hierarchical classification tree

To do so, we deploy and test five supervised classifiers: (1) naïve Bayes; (2) logistic regression; (3) support vector machines; (4) random forests; and (5) decision trees. Our selection of classifiers includes both linear and non-linear models[24], following prior benchmarking studies in data mining literature (e.g., Baesens et al., 2003; Loterman et al., 2012; Martens et al., 2007; Sebastiani, 2002; Van Gestel et al., 2004). Traditionally, researchers benchmark advanced machine learning models against linear techniques to conduct inference (Amani & Fadlalla, 2017). Naïve Bayes and logistic regression (also known as logit) are simple linear classifiers that perform well as common baseline models, against which more advanced classifiers can be tested (Amani & Fadlalla, 2017; Baesens et al., 2003; Henry, 2006; Martens et al., 2007; Provost & Fawcett, 2013). As a more advanced model, we investigate the performance of support vector machines, because they have been a state-of-the-art classifier and have proven

[24] Linear machine learning algorithms draw a linear decision boundary, while non-linear algorithms assume a non-linear relationship between classes (Huysmans, Dejaeger, Mues, Vanthienen, & Baesens, 2011; Ouyang et al., 2019).

to be a well-performing data mining technique (Martens et al., 2007; Moeyersoms et al., 2015; Vanhoeyveld et al., 2019). Support vector machines separate the data by constructing a hyperplane as a decision boundary in the feature space based on predefined classes, which can be used with both linear and non-linear kernel functions (i.e., radial basis function kernel, "*rbf*") (Martens et al., 2007; Purda & Skillicorn, 2015; Vanhoeyveld et al., 2019; Walker et al., 2020). Although prior research shows that support vector machines perform significantly more accurately than linear models (e.g., Goel et al., 2010; Van Gestel et al., 2003; Baesens et al., 2003 Loterman et al., 2012), their performance on attributions have not been widely investigated. As a non-linear technique, we use the random forest classifier in our analyses, as several benchmarking studies show that random forests provide the most accurate classification compared to other non-linear algorithms (Fernandez-Delgado et al., 2014; Guo et al., 2004; Lessmann et al., 2008; Moeyersoms et al., 2015). A random forest classifier is an ensemble method that generates multiple decision trees and aggregates the result of trees for a final class prediction (Breiman, 2001). Although non-linear techniques offer high predictive performance, their comprehensibility[25] is lower due to their complex mathematical models (Martens et al., 2007). Hence, we include an additional rule/tree-based classifier[26], which is considered the most comprehensible method (Martens et al., 2007; Moeyersoms et al., 2015) to provide a more robust comparison between models. Moreover, the classifiers we include in this study are shown to be effective in text classification and are widely used in prior research (Fernández et al., 2018, p. 3; Ignatow & Mihalcea, 2018, p. 62; Provost & Fawcett, 2013, p. 249).

3.1. Training dataset

We use a corpus of management commentary reports initially assembled and manually annotated by Aerts and Tarca (2010) as a dataset to train machine learning algorithms and

[25] The term "*comprehensibility*" refers to how well humans understand the classifier-induced classification.
[26] The specific decision tree algorithm we include in our analyses is "classification and regression tree" (also known as *CART in the literature*).

perform our tests. This corpus includes management commentary reports (MD&A or operating and financial review (OFR) reports) of listed companies from four common law countries (USA, Canada, UK, and Australia) where public accountability and related information disclosure is central to efficient capital markets. The companies in the corpus come from five industries (building materials, food processors, pharmaceuticals, biotechnology, and retail). The corpus consists of approximately 1.21 million words and comprises a total of 4,585 annotated attributional statements across 172 companies for the 2003 financial year (Table 1).

Table 1: Sample selection - number of companies by industry and country

	Biotech	Building mat.	Food process.	Pharma.	Retail.
Australia	7	9	7	7	8
Canada	10	5	8	8	5
UK	10	7	10	7	13
USA	13	4	11	12	11

Knowing that institutional setting may affect a company's attributional style, the multi-country dimension of the corpus allows to include more variety and intensity of attribution patterns in management commentary reports. As prior research suggested that company size and industry membership are associated with disclosure (Cole & Jones, 2004), the corpus controls for industry effects and company size. Industry is controlled by including companies of only five industries (cf. supra). Moreover, in each industry group and country, companies were selected based on relative size (measured by market capitalization)[27].

Although the management reports included in the corpus date from a while back, the corpus has the advantage of including, annotating, and corroborating a wide range of fine-grained attributional style properties which are relatively time-invariant by nature although

[27] For further details on selection details and population representation we refer to Aerts and Tarca (2010).

their prominence is affected by institutional context. In this vein, Aerts and Tarca (2010) show that differences in institutional environment and associated regulatory and litigation risks significantly affect the attributional properties of explanatory statements. Country differences relate to intensity of argument, presentational tendencies, preferences for formal language use and relative importance of tactical causal shading of explained outcomes through the use of entitlements, enhancements, excuses, justifications, and causality denials. Acknowledging this broad repertoire of attributional shading makes this corpus and dataset unique to explore the feasibility of automated attributional behavior measurement.

Using this dataset, we implement our three-level hierarchical classification approach. To construct our training/test data in the first level, we randomly select and manually classify 842 (421 in each class) intra-sentential and 428 (214 in each class) inter-sentential performance-related statements and non-performance-related statements from the MD&A reports. Non-performance-related statements provide information such as definitions of terms used in disclosures or company background that are not related to financial performance. Performance-related statements elaborate on a company's financial performance (with and without explanations). This initial set of statements from management commentary reports is used to train and test machine learning algorithms in a binary setting to identify performance-related statements in management commentary reports. Then, at the second level, we randomly select and manually identify 778 (389 in each class) intra-sentential and 428 (214 in each class) inter-sentential performance-related statements with attributions and performance-related statements without attributions. We use these observations to train and test the machine learning algorithms in a binary setting to detect performance-related statements with attributions, i.e., causality extraction, following prior literature (Zhao et al., 2016). The attributional statements used at the third classification level were initially identified by two independent researchers (coder 1 and 2) and characteristics of the identified attributional statements were manually coded by two other independent researchers (coder 3 and 4) according to a detailed coding scheme[28].

Each attributional statement is decomposed into an explained effect section and one or more explanatory phrases. The explained effects are coded according to four characteristics: nature (revenues, expenses, income/earnings/profit), valence (positive, negative/neutral), qualitative/quantitative, and analytical level of the effect (division/product/geographic, company as a whole). The explanatory phrases are also coded based on four characteristics: direction of cause-effect relationship (same direction, opposite direction), qualitative/quantitative, nature of cause (causal explanation, accounting-technical explanation), and locus of causality (internal, external). The performance-related attributional statements consist of 2,545 intra-sentential and 2,040 inter-sentential statements. This, in total (for all levels), provides us with 7,061 statements to train our classifiers. Table 2 shows the total number of statements at all three levels.

[28] Inter-coder reliability of the coding process showed initial agreement between coders for identifying attributional statements of 91%, while the initial agreement between coders for the coding characteristics of the components of attributional statements was, on average 88% (Aerts & Tarca, 2010).

Table 2: Total number of statements at all levels

		Intra-sentential	Inter-sentential
		n	*n*
Level 1: Performance x Non-performance:			
	1. Performance	421	214
	2. Non-performance	421	214
	Total	842	428
Level 2: With attributions x Without attributions:			
	1. With attributions	389	214
	2. Without attributions	389	214
	Total	778	428
Level 3: Characteristics:			
Outcome/effect (A's)			
A1 - Nature of effect	1. Revenue	857	659
	2. Expense	557	573
	3. Income/earnings	1,131	808
	Total	2,545	2,040
A2 - Valence of the effect	1. Positive	1,538	1,202
	2. Negative/unchanged	1,007	838
	Total	2,545	2,040

A3 - Effect is quantitative or qualitative	1. Quantitative	1,623	1,716
	2. Qualitative	922	324
	Total	2,545	2,040
A4 - Level of effect	1. Division/product/geographic	1,436	816
	2. Company as a whole	1,109	1,224
	Total	2,545	2,040
Cause/reasons (B's)			
B1 - Direction of cause-effect relationship	1. Same direction	2,253	1,758
	2. Opposite direction	292	282
	Total	2,545	2,040
B2 - Cause/reason is quantitative or qualitative	1. Quantitative	448	506
	2. Qualitative	2,097	1,534
	Total	2,545	2,040
B3 - Nature of cause/reason	1. Causal explanation	1,746	1,175
	2. Accounting technical explanation	799	865
	Total	2,545	2,040
B4 - Locus of causality	1. Internal	1,947	1,646
	2. External	598	394
	Total	2,545	2,040

As Table 2 shows, for the third level, the distribution between classes is skewed. Table 3 presents the imbalance ratio for each class of characteristics used at level 3. This can pose a challenge for classifiers because they tend to be biased towards the majority class (Chawla et al., 2002; Krawczyk, 2016; Ling & Li, 1998). A dataset is imbalanced when there is a significant or extreme disproportion among the number of examples in classes (Fernández et al., 2018, p. 19). A suggested solution for this kind of problem is manually under-sampling the major class and/or over-sampling the minor class (Batista et al., 2004; Chawla et al., 2002; Ling & Li, 1998). The former methodology randomly removes observations from the majority class until the total number of observations in the majority class is equal to the minority class, and the latter randomly uses an observation multiple times to train the algorithms (Batista et al., 2004). Chawla et al. (2002) introduce a more advanced over-sampling method for imbalanced distributions by creating synthetic examples rather than by over-sampling with replacements, called SMOTE ("*synthetic minority over-sampling technique*").

In this paper, we use SMOTE to over-sample the minority classes when the imbalance ratio is higher than 4.5:1[29]. In our dataset, we use SMOTE on four categories: Intra-sentential B1, intra-sentential B2, inter-sentential A3, and inter-sentential B1. We implement SMOTE on our training datasets to set the imbalance ratio to 2:1 by using "imblearn" Python library for the sampling technique. This is used to oversample the minority classes until the imbalance ratio is 2:1.

[29] The imbalance ratio is defined as the sample size of the largest majority class examples divided by the sample size of the smallest minority class (Fernández et al., 2018, p. 20).

Table 3: Imbalance ratio (IR) in each class of characteristics used in Level 3

Characteristics	Classes	Intra-sentential	IR	Inter-sentential	IR
A1 - Nature of effect	1. Revenue	857		659	
	2. Expense	557	2.0:1	573	1.4:1
	3. Income/earnings	1,131		808	
A2 - Valence of the effect	1. Positive	1,538	1.5:1	1,202	1.4:1
	2. Negative/unchanged	1,007		838	
A3 - Effect is quantitative or qualitative	1. Quantitative	1,623	1.8:1	1,716	5.3:1*
	2. Qualitative	922		324	
A4 - Level of effect	1. Division/product/geographic	1,436	1.3:1	816	1.5:1
	2. Company as a whole	1,109		1,224	
B1 - Direction of cause-effect relationship	1. Same direction	2,253	7.7:1*	1,758	6.2:1*
	2. Opposite direction	292		282	
B2 - Cause/reason is quantitative or qualitative	1. Quantitative	448	4.6:1*	506	3.0:1
	2. Qualitative	2,097		1,534	
B3 - Nature of cause/reason	1. Causal explanation	1,746	2.1:1	1,175	1.3:1
	2. Accounting technical explanation	799		865	
B4 - Locus of causality	1. Internal	1,947	3.2:1	1,646	4.2:1
	2. External	598		394	

3.2. Data preparation and text pre-processing

Sentence segmentation (one of the data pre-processing steps) and coding can be easily implemented to attributional statements that are embedded in one sentence through conjunctions. However, attributional statements can also span multiple sentences (Kruengkrai et al., 2017; Oh et al., 2013; Zhao et al., 2016), so that sentence-level analysis would not be robust (*example statement 2, cf. Section 2.1.2. Attribution*). Hence, we train machine learning algorithms at both sentence-level and multiple sentence-level separately, to capture both intra-sentential and inter-sentential features of attributions. At both intra-sentential and inter-sentential levels, it was necessary to adjust the coded attributional statements. In the dataset, some attributional statements combine one effect with multiple reasons or causes. For example, there are statements where one explained effect is attributed to several antecedents with commas and these antecedents may have different characteristics. So, each antecedent-consequence relationship was treated as a separate attributional statement. Below, a sample coded statement is provided:

> Example statement 3: *"Statutory operating profits for the year ended 31 March 2003 were £24.5 million. This was an increase of £6.6 million (37.2%) on the previous year. The newly acquired bonmarché business contributed £11.7 million to this figure and goodwill and exceptionals reduced it by £2.3 million, with the balance attributable to existing operations" (*The Peacock Group Plc. Annual report and accounts, 2003, p. 14).

In *example statement 3*, the explained outcome is that the company increased statutory operating profits by 37.2%. Different antecedent factors are mentioned to frame the performance outcome: (1) *The newly acquired 'bonmarché business' that contributed £11.7 million*; (2) *'goodwill and exceptionals' reducing performance by £2.3 million*, and (3) *with the balance 'attributable to existing operations*'. So, one performance outcome is linked with three antecedent (or causal) factors. One of the three antecedents has a negative effect on the outcome, while the other two contribute positively, meaning that antecedents have a different ‘direction of cause-effect relationship’ (one of the coded characteristics of the antecedents). In our analyses, we normalized the attributional statements by treating each combination of effect and antecedent as an individual attributional statement and treating the individual combination of effect and antecedent as unit of analysis. This is necessary for training machine learning algorithms so that they can understand patterns in the data. Table 4 shows an example of the detailed scoring and normalization of a combined attributional phrase.

Table 4: Example Coded Statement

Food Services operating profit has increased by 17% to £35.1 million reflecting a good performance in the household business and the improved trading conditions in commodity ingredients during the year.	
Outcome/effect *Food Services operating profit has increased by 17% to £35.1 million.* A1 - Nature of effect: Profit A2 - Valence of the effect: Positive A3 - Quantitative or Qualitative:Quantitative A4 - Level of effect: Division	
Cause 1 *reflecting a good performance in the household business.*	Cause 2 *and the improved trading conditions in commodity ingredients during the year.*
B1 - Direction of cause: Same direction B2 - Quantitative or qualitative: Qualitative B3 - Nature of cause: Causal explanation B4 - Locus of causality: Internal	B1 - Direction of cause: Same direction B2 - Quantitative or qualitative: Qualitative B3 - Nature of cause: Causal explanation B4 - Locus of causality: External
Phrase 1 *Food Services operating profit has increased by 17% to £35.1 million reflecting a good performance in the household business.*	**Phrase 2** *Food Services operating profit has increased by 17% to £35.1 million reflecting the improved trading conditions in commodity ingredients during the year.*

Before working on machine learning algorithms, several text pre-processing steps are required, such as sentence segmentation, tokenizing, stemming, and removing stop words. We use Python 3 on Jupyter Notebook for NLP and machine learning tasks. In Python, toolkits such as Spacy and natural language toolkit (NLTK) are used for NLP. Using Spacy, we first perform sentence segmentation for intra-sentential attributions. For inter-sentential attributions, we separate each adjacent two sentences. It has been reported that in 77% of multiple sentence statements, the first clause comes immediately after the first sentence to establish a connection (Prasad et al., 2008). So, for inter-sentential attributions, the explanatory features can be expected to be extracted from consecutive sentences (Walker et al., 2020). Then, we remove stop words and tokenize performance-related attributional statements. We use NLTK's Porter stemmer to remove all suffixes and prefixes in each word. We also employ the bigram identification in Python, so we train the algorithms by controlling the combination of two adjacent words. For the term weighting method, we use TF-IDF, which is a common value representation method for terms (Bach et al., 2019; Liu et al., 2017; Loughran & McDonald, 2011; Provost & Fawcett, 2013, p. 256).

3.3. Algorithm construction

For machine learning, we use the scikit-learn toolkit in Python. The scikit-learn toolkit is *"a Python module integrating a wide range of state-of-the-art machine learning algorithms for medium-scale supervised and unsupervised problems"* and is widely used in machine learning research (Pedregosa et al., 2011).

As all classifiers use different parameters for calculation, it is necessary to tune the parameters of the machine learning models, so that the model fits the data as good as possible (Provost & Fawcett, 2013, p. 81). To find hyper-parameters for optimal performance, we use GridSearchCV of the Scikit-learn toolkit in Python, which selects the parameters on a specified parameter grid, fits it to the algorithm, and reveals the maximum performance for each classifier (Pedregosa et al., 2011). In the *K*-fold cross-validation test, the training and test sets are randomly divided into *K* number of parts, and algorithms are run *K* times. One of the *K* parts is used as the testing data, while the others are used as the training data. This is a bootstrapping technique to increase the validity of results (Humpherys et al., 2011). For our grid-search technique, we use the *10*-fold cross-validation method to find out the optimal parameter combination of classifiers. For naïve Bayes classifier parameters, we control different *alpha*[30] values, while other parameters are taken as default values. For logistic regression classifier parameters, we control *random state*[31] and *solver*[32]. For support vector machines' parameters, we control the *penalty parameter C of the error term*[33] and *the kernel*[34]. For random forests, we control *the number of trees in the forest*[35] and *the maximum depth of the tree*[36]. Finally, for decision trees we control *the maximum depth of the tree* and *the minimum number of samples required to be at a leaf node.*[37]

With the selected parameters, we train and test the five classifiers with the pre-processed attributional statements. We conduct the analysis separately for intra- and inter-sentential attributional statements using a 15% train-test split for all levels. To control the effectiveness of machine learning methods, we compare accuracy, F1-score, and wall time of the grid search process. Accuracy is a common performance evaluation metric and assesses the overall effectiveness of the algorithm (Sokolova et al., 2006). It is calculated by the total number of statements correctly classified, divided by the total number of statements analyzed. Because accuracy does not distinguish between the numbers of correctly classified examples of different classes (Fernández et al., 2018, p. 22), we measure the performance of the classifiers using F1-score, as well. Van Rijsbergen (1979) defines the F1-score as the harmonic mean of precision and recall. F1-score is *"a composite measure which benefits algorithms with*

[30] Alpha: Additive (Laplace/Lidstone) smoothing parameter. Controlled for 0.01, 0.1 and 1.0 (default: 1.0).
[31] Random state: The seed of the pseudo random number generator to use when shuffling the data. Controlled for "none", 42, 100.
[32] Solver: Controlled for 'lbfgs', 'newton-cg', 'sag' and 'saga' (default: lbfgs).
[33] C: Regularization parameter. Controlled for 1, 8, 12 and 64 (default: 1).
[34] Kernel: Specifies the kernel type to be used in the algorithm. Controlled for 'linear' and 'rbf' (default: 'rbf').
[35] N_estimator: The number of trees in the forest. Controlled for 10, 100 and 1000 (default: 100).
[36] Max_depth: The maximum depth of the tree. Controlled for "none", 5, 10, 100, 1000 (default: "none").
[37] Min_samples_leaf: The minimum number samples required to be at a leaf node. Controlled for .04, .06, .08, 1 (default: 1).

higher specificity" (Purda & Skillicorn, 2015; Sokolova et al., 2006). In text classification, the metrics precision and recall are often used (Provost & Fawcett, 2013, p. 271). Precision is the ratio of relevant observations retrieved to the total number of observations retrieved, and recall is the ratio of the number of relevant observations retrieved to the total number of relevant observations (Formula 1) (Yang et al., 2018).

$$F1_{Score} = 2 * \frac{precision * recall}{precision + recall} \quad (1)$$

Moreover, we include the wall time of each grid search process for each classification to measure the performance of classifiers in terms of time. Wall time is the actual time that is needed to run the program from the start to the end.

4. Results

4.1. Performance evaluation of the classifiers

We follow a top-down strategy to classify attributional features of performance-related statements (Figure 1)[38]. For hierarchical classification tasks, the overall classification tasks are structured with subclasses and superclasses (Nakano et al., 2017). In the top-down approach, one or more classifiers are trained for each level in the hierarchy (Costa et al., 2007; Kiritchenko et al., 2005).

We plot the predictive performance of our models using the Matplotlib library on Python. Figure 2 illustrates the predictive performance of classifiers on all levels for intra- and inter-sentential settings. It is apparent that our decision tree classifier is the least accurate model for all levels and settings. This result is in line with prior research as a rule/tree-based classifier provides more comprehensibility and limited predictive performance (Moeyersoms et al.,

[38] We conduct our analysis using a system with IntelCore i7-8850H processor, 16.0 GB installed memory (RAM). The Python codes of the analyses are available in the following repository: https://github.com/anilberkin/nlp-ml/blob/main/ML_Comparison.ipynb

2015). However, the most accurate classifier varies depending on intra- and inter-sentential settings, as well as the specific level being considered.

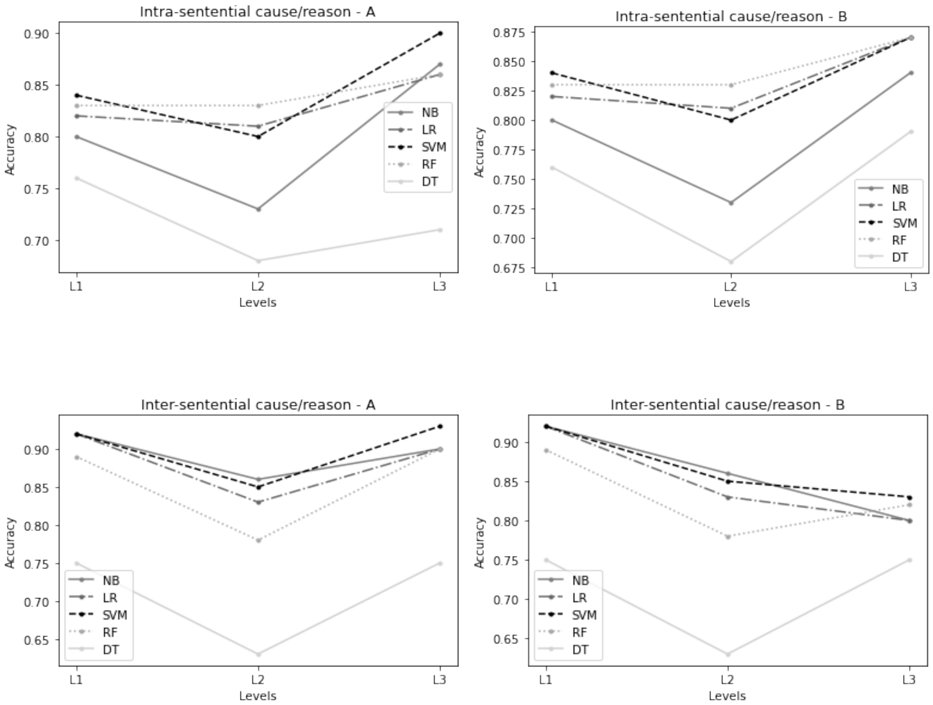

Figure 2: Accuracy results at levels

**NB: Naïve Bayes classifier, LR: Logistic regression, SVM: Support vector machines, RF: Random forest, DT: Decision trees*

The first level is to extract performance-related statements and separate them from all other information disclosed in narrative reports, i.e., identification. The one-versus-rest binary (performance-related statements and non-performance-related statements) classification results show that all five machine learning techniques have an F1-score and accuracy that equals at least 75%, with support vector machines (trained with the "*rbf*" kernel as a non-linear model) performing as the most accurate classifier (see Table 5). The results also indicate that classifiers perform better with multiple sentences (inter-sentential) than a single sentence (intra-sentential) in terms of classifying whether statements are performance-related or not.

At the second level, we focus on detecting performance-related statements with attributions, in which managerial explanations on performance outcomes are provided. We train machine learning algorithms in a binary setting, i.e., performance-related statements with and without attributions, and test the performance of classifiers. For the second level, the performance of classifiers is lower compared to the first level in both intra- and inter-sentential settings. This shows that it is more challenging to classify performance-related statements based on whether or not the identified performance-related statements include managerial explanations compared to distinguishing performance-related statements from non-performance-related statements. The finding is in line with prior research focusing on causality mining, or causality extraction (Kruengkrai et al., 2017; Oh et al., 2013; Zhao et al., 2016). The highest F1-score and accuracy (being 83%) are noted for the random forest classifier for the intra-sentential setting. Interestingly, the naïve Bayes classifier outperforms the other classifiers for the inter-sentential setting (F1-score and accuracy of 86%), while the random forest classifier ranks as the last but one (and after decision trees) based on its performance for the inter-sentential setting (with 78% accuracy and 79% F1-score).

Table 5: Results of Level 1 and Level 2

		Intra-sentential			Inter-sentential		
	Classifier	Acc.	F1	Time	Acc.	F1	Time
Level 1	NB	.80	.80	3"	.92	.92	1"
	LR	.82	.82	9"	.92	.92	7"
	SVM	.84	.84	19"	.92	.92	10"
	RF	.83	.83	1'25"	.89	.89	1'00"
	DT	.76	.75	4"	.75	.75	3"
Level 2	NB	.73	.73	2"	.86	.86	1"
	LR	.81	.81	7"	.83	.83	6"
	SVM	.80	.80	14"	.85	.85	9"
	RF	.83	.83	1'24"	.78	.79	1'06"
	DT	.68	.68	4"	.63	.63	3"

At the third level, we analyze the performance of classifiers in classifying performance-related attributional statements based on content and style-related characteristics. In Table 6, the performance evaluation metrics for the machine learning algorithms are shown for each characteristic. Depending on the content and style characteristics, the classification accuracy and F1-score vary in the range of 58%-94%. The results clearly show that our support vector machines largely outperform other classifiers at the third level for both outcome/effect and cause/reason characteristics in intra- and inter-sentential settings. The lowest accuracy and F1-score on both intra- and inter-sentential settings are observed for B3 - nature of cause, i.e., whether explanations are accounting-technical or based on causal reasoning.

Table 6: Results of Level 3

		Intra-sentential			Inter-sentential		
Characteristics	Classifier	Acc.	F1	Time	Acc.	F1	Time
A1 – Nature of effect	NB	.88	.88	5"	.91	.91	7"
	LR	.91	.91	44"	.92	.92	28"
	SVM	.93	.93	3'11"	.94	.94	1'41"
	RF	.91	.91	5'47"	.94	.94	2'56"
	DT	.77	.77	9"	.77	.77	10"
A2 – Valence of the effect	NB	.87	.87	6"	.90	.90	4"
	LR	.85	.85	36"	.87	.87	21"
	SVM	.89	.89	3'26"	.92	.92	1'42"
	RF	.85	.85	6'08"	.89	.89	2'54"
	DT	.68	.67	9"	.73	.72	11"
A3 – Effect is quantitative or qualitative	NB	.84	.84	6"	.92	.92	8"
	LR	.84	.83	38"	.93	.92	32"
	SVM	.88	.88	3'36"	.91	.91	1'15"
	RF	.85	.84	6'29"	.92	.93	2'32"
	DT	.73	.72	9"	.82	.82	17"
A4 – Level of effect	NB	.88	.88	3"	.89	.89	6"
	LR	.85	.84	19"	.90	.90	22"
	SVM	.89	.89	1'53"	.93	.93	1'40"
	RF	.84	.84	3'51"	.88	.88	2'57"

	DT	.64	.62	8"	.69	.67	9"
B1 – Direction of cause-effect relationship	NB	.86	.86	4"	.82	.81	4"
	LR	.94	.94	20"	.85	.83	23"
	SVM	.94	.94	1'14"	.85	.84	1'18"
	RF	.93	.92	2'33"	.88	.85	2'42"
	DT	.92	.92	7"	.81	.80	8"
B2 – Cause/reason is quantitative or qualitative	NB	.88	.87	3"	.82	.79	5"
	LR	.89	.89	21"	.79	.72	21"
	SVM	.88	.87	1'36"	.85	.83	1'12"
	RF	.90	.89	2'50"	.81	.75	2'40"
	DT	.73	.76	7"	.81	.78	9"
B3 – Nature of cause/reason	NB	.79	.79	6"	.74	.74	4"
	LR	.81	.80	41"	.74	.74	23"
	SVM	.82	.82	3'48"	.76	.76	2'43"
	RF	.80	.79	7'23"	.74	.74	4'14"
	DT	.71	.67	8"	.59	.58	10"
B4 – Locus of causality	NB	.84	.83	3"	.81	.79	6'
	LR	.83	.80	21"	.83	.79	39'
	SVM	.85	.83	1'39"	.85	.83	2'13"
	RF	.83	.82	3'30"	.83	.81	4'34"
	DT	.78	.72	9"	.78	.68	9"

Finally, the results show that running the hyper-parameter selection tool takes longer for random forests compared to the other classifiers. Although the naïve Bayes algorithm is not the classifier with best performance, it may still be a good alternative to use for large datasets, due to its low computational cost (cf. wall time).

In order to demonstrate the performance of the classifiers in more detail, confusion matrices, i.e., true (false) positives (negatives), are provided for each level in Table 7, which is used to calculate F1-scores (i.e., precision and recall). We additionally use the confusion matrices to evaluate the hierarchical model.

Table 7: Confusion Matrices/Level

	Classifiers	Intra-sentential TP	FP	TN	FN	Inter-sentential TP	FP	TN	FN
Level 1									
Performance x Non-performance	NB	52	13	50	12	30	1	29	4
	LR	49	8	55	15	30	1	29	4
	SVM	51	7	56	13	29	1	30	4
	RF	49	6	57	15	29	1	28	5
	DT	42	9	54	22	24	6	24	10
Level 2									
With attributions x Without attributions	NB	43	12	42	20	26	6	30	3
	LR	51	10	44	12	26	8	28	3
	SVM	51	11	43	12	26	7	29	3
	RF	56	13	41	7	23	8	28	6
	DT	45	19	35	18	21	16	20	8
Level 3 - Characteristics									
A1 – Nature of effect	NB	113	16	242	16	96	9	200	9
	LR	117	11	246	11	96	8	200	8
	SVM	120	9	249	9	99	6	203	6
	RF	117	11	246	11	98	4	203	6
	DT	100	29	228	29	81	24	186	24
A2 – Valence of the effect	NB	126	23	212	26	127	17	156	14
	LR	118	23	212	34	119	18	155	22
	SVM	133	23	212	19	128	12	161	13

	RF	114	21	214	38	118	10	163	23
	DT	78	51	184	74	87	32	141	54
A3 – Effect is quantitative or qualitative	NB	98	18	226	42	31	13	250	13
	LR	94	16	228	46	26	5	258	18
	SVM	113	18	226	27	27	10	253	17
	RF	95	13	231	45	24	0	263	20
	DT	75	39	205	65	17	27	236	27
A4 – Level of effect	NB	133	18	207	28	173	20	106	14
	LR	120	18	207	41	176	21	105	11
	SVM	135	18	207	26	176	11	115	11
	RF	123	25	200	38	171	20	106	16
	DT	84	62	163	77	148	69	57	39
B1 – Direction of cause-effect relationship	NB	25	27	306	27	17	18	241	37
	LR	36	6	327	16	20	14	245	34
	SVM	40	10	323	12	22	14	245	32
	RF	30	6	327	22	19	4	255	35
	DT	31	8	325	21	16	21	238	38
B2 – Cause/reason is quantitative or qualitative	NB	296	26	23	19	201	43	22	7
	LR	302	26	23	13	207	57	8	1
	SVM	300	28	21	15	200	34	31	8
	RF	310	31	18	5	207	52	13	1
	DT	248	31	18	67	199	43	22	9
B3 – Nature of cause/reason	NB	66	28	235	51	99	49	126	31

	LR	64	18	245	53		92	41	134	38
	SVM	71	21	242	46		99	41	134	31
	RF	58	16	247	59		90	38	137	40
	DT	31	24	239	86		57	52	123	73
	NB	44	13	281	48		23	12	232	47
B4 – Locus of causality	LR	31	6	288	61		19	3	241	51
	SVM	42	6	288	50		28	5	239	42
	RF	39	7	287	53		23	5	239	47
	DT	14	7	287	78		19	11	230	54

The evaluation of the performance of classifiers at each level does not cover the performance of classifiers for a hierarchical classification task, which is more complex (Costa et al., 2007). Because of the relations to other classes in the hierarchy, e.g., ancestors or descendants, the commonly used measures such as accuracy, F-measure, precision, or recall are not appropriate (Kosmopoulos et al., 2014).

In our model, we employ hierarchy-based evaluation metrics, i.e., hierarchical F1-score, to measure the overall performance of the three-level hierarchical classification task. This is an evaluation metric based on hierarchical precision and hierarchical recall (Cerri et al., 2016; Costa et al., 2007; Kiritchenko et al., 2004; Nakano et al., 2017). Formulas (2, 3, 4) are used to compute the hierarchical F1-score, where hP and hR stands for hierarchical precision and hierarchical recall and C_i and Z_i correspond to the set of true and predicted classes for an instance *i,* respectively:

$$h\mathrm{P} = \frac{\sum_i |Z_i \cap C_i|}{\sum_i |Z_i|} \quad (2)$$

$$h\mathrm{R} = \frac{\sum_i |Z_i \cap C_i|}{\sum_i |C_i|} \quad (3)$$

$$hF_{score} = \frac{2 * h\mathrm{P} * h\mathrm{R}}{h\mathrm{P} + h\mathrm{R}} \quad (4)$$

To calculate the hierarchical F1-score of an instance i, we employ a gold standard, i.e., human-annotated, external data. We use data provided by El Haj et al. (2016), which consists of 3,491 coded intra-sentential performance-related statements based on *locus of causality*, i.e., internal or external attributions, from 500 preliminary earnings announcements issued between 2010 and 2012 by firms listed on the London Stock Exchange. We randomly select 100 attributional statements, *i,* from the dataset and use this as a holdout sample to see whether our trained classifiers can predict the correct classes, i.e., B4 - locus of causality: external or internal. We run three levels, as in a top-down hierarchical setting, on our best-performed classifier, support vector machines[39], and measure the hierarchical F1-score for the classification based on locus of causality using the ancestor classes, i.e., level 1 and level 2. The hierarchical F1-score for locus of causality is 61%. This is lower than non-hierarchical results because the hierarchical measure punishes errors at higher levels more heavily (Kiritchenko et al., 2005). Due to unavailability of needed detail in external gold standard datasets, it is not possible to evaluate the hierarchical F1-score for all content and style-related attributional characteristics. Thus, we used locus of causality as a proxy for the third level. However, we can generalize that 84%, 80%, and 83% (for the first, second, and third levels, respectively) F1-scores lead to hierarchical F1-score varying between 60% - 65%.

4.2. Supplementary analyses

The utility of a predictive model depends on the model's ability to maintain its accuracy when different samples are used as input values to the model (Terrin et al., 2003). To assess this, we conduct a supplementary external validity test on support vector machines using an alternative dataset consisting of chairman's statements. Chairman's statements are highly discretionary and subject to IM because managers' predisposition to associate themselves with corporate financial results is linked to the firm's underlying financial performance (Clatworthy & Jones, 2006). Hence, they provide a good ground to test the validity of support vector machines. Using Bureau Van Dijk's Orbis database, we identified listed firms operating in the UK for fiscal year 2018. Next, based on change in return on assets (ROA) between 2017 and 2018, we selected firms in the top and bottom quartiles to obtain a sample including both the best and worst performing firms. We exclude observations with extreme changes in ROA (i.e., more than 100% in- or decrease), firms that operate in the financial sector (NACE code: 64, 65, 66), firms that operate in "other activities" (NACE code: S), and firms that do not have a

[39] Support vector machines give the best accuracy among other classifiers in our internal validity tests (average accuracy is 0.88 for three levels, with best-fit hyper-parameters: C=8, gamma=1, kernel='rbf').

chairman's statement in their annual report. Doing so, we end up with sample of 212 chairman's statements (collected through firms' websites and the website of the London Stock Exchange), resulting in a final sample of 9,974 sentences in total. Having a sample composed of both the best and worst performing firms ensures a variety of performance-related attributions in our dataset.

We use the sentences as an input for our trained support vector machines with best-fit hyper-parameters. We, then, randomly select 100 intra-sentential performance-related attributional statements from the chairman's statements, manually annotate their outcome/effect and cause/reason according to their content and style and track their classification. For each chairman's statement, our classifier follows level 1 (identification of performance-related statements), level 2 (detection of performance-related statements with attributions), then level 3 (classification of the identified performance-related statements with attributions based on the textual characteristics). For the first level, we expect our classifier to extract the sentences that we randomly selected. In total, our classifier extracts 2,883 performance-related statements for the first level. 86 out of the 100 randomly selected performance-related statements with attributions are classified as performance-related statements (implying a 14% error rate for the first level). For the external validity of the second level, we run our algorithm to classify the 86 performance-related statements based on whether they contain attributions or not. Out of the 86 correctly classified performance-related statements (with attributions), 23 were classified as without attributions. 23 misclassifications imply a 27% error rate (i.e., 23/86), which is in line with our findings. For the third level, on average, content and style-related characteristics of performance-related attributional statements can be classified with a 20% error rate. Based on this, the error rates *(1-accuracy)* for the hierarchical classification are shown in Table 8.

Table 8: External validity results – Error rates

Level 1 - Performance/Nonperformance	
Performance x Non-performance	.14
Level 2	
With attribution x Without attribution	.27
Level 3 - Characteristics	
A1 – Nature of the effect	.17
A2 - Valence of the effect	.14
A3 - Effect is quantitative or qualitative	.19
A4 - Level of effect	.34
B1 - Direction of cause-effect relationship	.12
B2 - Cause/reason is quantitative or qualitative	.18
B3 - Nature of cause/reason	.23
B4 - Locus of causality of cause/reason	.19

External validity results can be lower than internal accuracies as using a different dataset may result in a decrease in accuracy (Terrin et al., 2003). This is expected, because data characteristics, variable definitions, or data collection methods may change (Terrin et al., 2003). Moreover, external validity results also indicate that in a hierarchical setting 48[40] out of 100 statements are misclassified at the end of the hierarchy, where most of the misclassifications were found in the second level (27%), which is in line with prior research. Our results provide evidence that automated causality extraction (level 2) is the weaker phase, due to the multiple structures of causal inference (Kruengkrai et al., 2017; Oh et al., 2013; Walker et al., 2020; Zhao et al., 2016). As causality mining (second level in the hierarchical

[40] 14/100 of statements were misclassified in the first level, 23/86 of statements were misclassified in the second level and 11/63 were misclassified in the third level.

approach) remains the most challenging step with lower accuracy in the machine learning hierarchy, the errors of the second level are carried to the third level, which leads to a significant decline in overall performance of the hierarchical approach.

Alternatively, the LIWC causal word-list (Tausczik & Pennebaker, 2009) and the Coh-Metrix' causal connectives list (Graesser et al., 2011) provide a dictionary approach to detect attributions. To test the efficiency of the dictionary approach on causality extraction and compare the performance with machine learning models, we use the same 100 performance-related statements with attributions as input for LIWC (causality dictionary) and the Coh-Metrix (causal connectives) and compare their result with the machine learning algorithm on the second level. The results show that LIWC detects causality in 48/100 statements (i.e., the causal scores of the statements are higher than zero) and the Coh-Metrix captures causal connectives in 69/100 statements (i.e., causal connectives scores "*CNCCaus*" of the statements are higher than zero), while our support vector machines classify 76/100 of the statements correctly as "with attribution", which is more accurate than the Coh-Metrix classification. Manually investigating the outcomes of LIWC and the Coh-Metrix reveals that statements including attribution do not necessarily have to include causal words or phrases, that can be captured through LIWC and the Coh-Metrix. Below three example statements are given that were misclassified by both LIWC and the Coh-Metrix, but were accurately classified as "with attribution" by our machine learning algorithm:

> *"The majority of our markets performed well, demonstrating the values of a balanced portfolio, with notably strong performances in Automative, Electronics and Energy" (*Victrex plc., 2018)

> *"In North America, the US also had a record year, growing 25%, with our strategy of diversification, both in terms of location and discipline, delivering notable performances from our regional offices in Boston, Chicago, Houston and Los Angeles."* (Pagegroup plc, 2018)

> *"Group revenue at £367.5m was 11.0% ahead of 2017 reflecting in part the acquisition of Bison business completed in September 2017."* (Forterra plc., 2018)

As our comparison outcomes show, none of the automatized causality extraction methods provide as accurate results as manual coding, although implementing machine learning to detect attributions can provide significantly more robust classification than LIWC and the Coh-Metrix. Taken together, these results provide important insights into the feasibility of using machine learning algorithms for the classification of performance-related attributional statements. This section has reviewed the external validity performance of our best-performing machine learning algorithm, support vector machines, and compared their performance with traditional dictionary methods. The following section will present conclusions, discuss limitations and expand on future work.

5. Discussion and conclusion

The aim of the present study was to examine the feasibility of machine learning algorithms for the classification of performance-related attributional statements. As van Atteveldt et al. (2021) discuss, both traditional content analysis and machine learning methods have their advantages and disadvantages. Whether traditional content analysis or machine learning suits best for a study depends on the research question and the analysis of interest (El-Haj et al., 2019a). Nevertheless, the vast majority of studies on attributional IM have used a manual text analysis approach due to the complexity of managerial explanations (El-Haj et al., 2019a). Overall, our results show that machine learning algorithms may offer an alternative methodological approach for the analysis of attributional statements in narrative reports. Compared to traditional text analysis, machine learning can provide a fairly accurate methodology for large-scale identification and coding of attributions in a fast way, despite their complicated nature. Yet, manual content analysis remains unavoidable for automated classification due to the necessity of a training dataset (El-Haj et al., 2019a; Fisher et al., 2016).

In applying a top-down hierarchical classification approach, we replicate the full scope of a manual coding process of attributions using machine learning algorithms. The hierarchy relates to identifying statements commenting on firm performance (level 1), to detecting the ones with explanations (level 2), and, lastly, to classifying explanations based on attributional features (level 3). This approach allows us to observe the performance of classifiers at different coding levels, as well as for the full coding process, and to identify levels that require further attention or, eventually, alternative treatment. Due to the different forms of causal inference, our findings support previous studies, showing automated causality extraction (level 2) is the weaker phase, which results in an overall decline in the performance of the hierarchical approach. Although the performance of machine learning methods for level 2 is slightly better than dictionary methods (76% accuracy compared to 69% and 48% in the supplementary analyses), manual identification of managerial explanations may provide a more robust solution. Hence, using a hybrid approach, where machine learning methods for level 1 and level 3 combined with manual identification of attributions at level 2, would increase the hierarchical F1-score significantly by about 20% (considering no misclassifications at level 2). Overall, our findings are in line with prior research corroborating that automated techniques do not always perform sufficiently accurate and need to be validated (van Atteveldt et al., 2021; Walker et al., 2020). More specifically, our study highlights that automatic causality extraction needs to be carefully validated.

In addition, for the third level in the classification hierarchy, our results show that machine learning algorithms perform better for certain attributional features (e.g., nature of effect) than others (e.g., nature of cause). While machine learning algorithms perform sufficiently well (90% accuracy) in classifying thematic differences (i.e., revenue, expense, income) in the attributional statements, classification based on the nature of inference (causal explanation versus accounting-technical explanation) is more challenging. Again, this relates to the diversity in causal inferences that were included in the training set. Further research could use our findings (results at level 3) to determine how to use and train machine learning algorithms. In this regard, it may be worthwhile to consider different training sets for classification of causal explanations versus accounting-technical explanations. This might also alleviate accuracy concerns at the causality extraction level (level 2).

Moreover, while prior research so far has mainly focused on intra-sentential attributions and largely neglected inter-sentential attributions (due to the difficulty of the information extraction process), we consider both in our analyses and our results reveal a slight increase in the performance of machine learnings algorithms for inter-sentential attributions.

Comparing the performance of five widely-used classifiers based on performance indicators for all levels and features, the decision tree classifier appears to be the least accurate classifier, while support vector machines (with an 'rbf' kernel) appear to be the most accurate classifier, although it slightly varies depending on the level of classification and attributional feature. This is in line with previous findings (Domingos & Pazzani, 1997; Martens, et al., 2007; Moeyersoms et al., 2015; Fernández et al., 2018, p. 124; Goel et al., 2010; Walker et al., 2020). Although decision trees and naïve Bayes generally provide more comprehensible results, their predictive performance tends to be lower. We show that although more advanced models, i.e., non-linear, are computationally more costly (i.e., substantial increase in wall time), they are better to use to obtain higher predictive performance.

In general, we expect that the findings regarding the automated attributional analysis that we investigate in this study, are directly useful to investigate both the informational and presentational dimensions of explanations in managerial communication. While informational dimensions focus on how attributional statements create new information both in the sense of providing information on actual causality and building cognitive legitimacy, presentational dimensions relate to IM intent and primarily focus on selectivity and bias in the content patterns of the attributional explanations. Attributional statements tend to define essential elements of the corporate performance environment for both internal and external audiences. They portray the normative and empirical bases on which to take measures or to judge ex-post the appropriateness of a firm's actions and achievements by explaining "why" and "how" of corporate outcomes (Aerts, 2005). Prior research highlights the importance of such performance explanations for investors' judgments (Koonce et al., 2011; Cianci & Kaplan, 2010). Moreover, analyzing attributional statements is also particularly useful when investigating corporate governance processes, as attributions are highly instrumental to the accountability mechanisms that are embedded in the institutional environment of public companies to keep companies answerable and responsible for past behavior (Aerts & Tarca, 2010; Brown et al., 2011).

In practice, however, managerial performance explanations are not always causality-based. Performance evaluation is permeated by accounting language with its inherent self-explanatory logic: it provides its own categories (e.g., expenses, revenue, segments) to explain other (more aggregated) categories (e.g., margins and earnings) and accounting phenomena (such as consolidation of entity figures) (Aerts, 2005). The meaning of such formal language explanations is rather ambiguous, as the concepts and relationships that they use to explain performance outcomes are essentially analytical and do not establish actual causality (Aerts, 1994; Hines, 1988). Nevertheless, they build on taken-for-granted and prominent ways of recording, analyzing and presenting firm performance. Being generally expected and conventional, accounting explanations are quite prominent in managerial performance explanation. To the extent that causal explanations are descriptive of actual causal mechanisms, they are probably more informative than accounting explanations by pointing to responsibility, motives and goals, or by altering the perceived valence of an outcome by offering a causal context (Koonce et al., 2011). They may, however, reveal proprietary information on the underlying business model and proprietary costs may considerably constrain the use of causal explanations. In addition, causal explanations are not litigation-proof (Aerts & Cheng, 2011). Causation inferences are frequently established and interpreted based on intuition and conviction rather than factual facts. Moreover, in the diagnosis phase of explanation, it is not uncommon to identify a number of interrelated intervening factors which have been helpful or constraining in bringing about one specific outcome. An exposed argument, then, comes down to a selection process to arrive at one or more convenient explanations, where choosing a cause from a potentially large number of relevant antecedents is likely to be purpose-specific and context-dependent (Aerts et al., 2013). In that regard, performance explanation often reflects a tension between informational validity and presentational adequacy.

The machine learning tools that we investigate in this study, accommodate the variety of attributional phenomena that are encountered in practice and allow to elaborate and examine an array of attributional features, from the intensity of argument, obvious self-presentational biases and preferences for formal language use to tactical causal shading of explained outcomes through the use of entitlements, enhancements, excuses, justifications and causality denials. The feasibility analysis we conduct may be helpful to address a diversity of informational and presentational research questions and settings where managerial performance communication (such as board meeting reports, management commentary and management presentations, MD&A, analyst reports, conference calls, blogs) is used as a key research instrument.

Our study has certain limitations. First, we face general machine-learning-related limitations. For supervised machine learning methods, manual coding as a first step is necessary, as machine learning methods cannot discover original features unless they learn them from a training dataset (Goel et al., 2010). Because publicly (and readily) available training datasets are extremely scarce, constructing a training dataset is still a very time-consuming process. Second, sample size is a limitation of our study. As sample size increases statistical power for classifiers (El-Haj et al., 2019a), our study especially suffers from scarcity of training data for specific features (although we used oversampling to mitigate the issue). Third, scarcity of external datasets limited our study as well. We were not able to conduct additional hierarchical F1-score tests due to the absence of gold standard datasets. Fourth, following prior literature, due to the complexity of attributional statements, we had to normalize attributional statements by treating each combination of effect and antecedent as an individual attributional statement and treat the individual combination of effect and antecedent as unit of analysis. More research on attributions with multiple causes/reasons is needed.

Reference List

Abrahamson, E., & Park, C. (1994). Concealment of Negative Organizational Outcomes: An Agency Theory Perspective. *Academy of Management Journal, 37*(5), 1302-1334.

Aerts, W. (1994). On the use of accounting logic as an explanatory category in narrative accounting disclosures. *Accounting, Organizations and Society, 19*(4-5), 337-353. doi:10.1016/0361-3682(94)90001-9

Aerts, W. (2005). Picking up the pieces: impression management in the retrospective attributional framing of accounting outcomes. *Accounting, Organizations and Society, 30*(6), 493-517. doi:10.1016/j.aos.2004.07.001

Aerts, W., & Cheng, P. (2011). Causal disclosures on earnings and earnings management in an IPO setting. *Journal of Accounting and Public Policy, 30*(5), 431-459. doi:10.1016/j.jaccpubpol.2011.03.006

Aerts, W., Cheng, P., & Tarca, A. (2013). Management's Earnings Justification and Earnings Management under Different Institutional Regimes. *Corporate Governance: An International Review*, 21(1), 93-115. doi:10.1111/corg.12014

Aerts, W., & Tarca, A. (2010). Financial performance explanations and institutional setting. *Accounting and Business Research, 40*(5), 421-450. doi:10.1080/00014788.2010.9995322

Aerts, W., & Yan, B. (2017). Rhetorical impression management in the letter to shareholders and institutional setting. *Accounting, Auditing & Accountability Journal, 30*(2), 404-432. doi:10.1108/aaaj-01-2015-1916

Amani, F. A., & Fadlalla, A. M. (2017). Data mining applications in accounting: A review of the literature and organizing framework. *International Journal of Accounting Information Systems, 24*, 32-58. doi:10.1016/j.accinf.2016.12.004

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of Internet stock message boards. *Journal of Finance, 59*(3), 1259-1294. doi:DOI 10.1111/j.1540-6261.2004.00662.x

Aureli, S. (2017). A comparison of content analysis usage and text mining in CSR corporate disclosure. *The International Journal of Digital Accounting Research, 17*, 1-32. doi:10.4192/1577-8517-v17_1

Bach, M. P., Krstić, Ž., Seljan, S., & Turulja, L. (2019). Text Mining for Big Data Analysis in Financial Sector: A Literature Review. *Sustainability, 11*(5). doi:10.3390/su11051277

Baesens, B., Setiono, R., Mues, C., & Vanthienen, J. (2003). Using Neural Network Rule Extraction and Decision Tables for Credit-Risk Evaluation. *Management Science, 49*(3), 312-329. doi:10.1287/mnsc.49.3.312.12739

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring. *The Journal of the Operational Research Society, 54*(6), 627-635. doi:10.1057/palgrave.jors.2601545

Baginski, S. P., Hassell, J. M., & Hillison, W. A. (2000). Voluntary Causal Disclosures: Tendencies and Capital Market Reaction. *Review of Quantitative Finance and Accounting, 15*(4), 371-389. doi:10.1023/A:1012002608615

Baginski, S. P., Hassell, J. M., & Kimbrough, M. D. (2004). Why Do Managers Explain Their Earnings Forecasts? *Journal of Accounting Research, 42*(1), 1-29. doi:10.1111/j.1475-679X.2004.00127.x

Baginski, S. P., Hassell, J. M., & Wieland, M. M. (2011). An examination of the effects of management earnings forecast form and explanations on financial analyst forecast revisions. *Advances in Accounting, 27*(1), 17-25. doi:10.1016/j.adiac.2011.01.005

Batista, G. E. A. P. A., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *SIGKDD Explor. Newsl., 6*(1), 20–29. doi:10.1145/1007730.1007735

Bickel, M. W. (2017). A new approach to semantic sustainability assessment: text mining via network analysis revealing transition patterns in German municipal climate action plans. *Energy, Sustainability and Society, 7*(1). doi:10.1186/s13705-017-0125-0

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research, 3*, 993-1022.

Boritz, J. E., Hayes, L., & Lim, J.-H. (2013). A content analysis of auditors' reports on IT internal control weaknesses: The comparative advantages of an automated approach to control weakness identification. *International Journal of Accounting Information Systems, 14*(2), 138-163. doi:10.1016/j.accinf.2011.11.002

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). *The Development and Psychmetric Propoerties of LIWC-22*. The University of Texas at Austin.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5-32. doi:10.1023/A:1010933404324

Brown, P., Beekes, W., & Verhoeven, P. (2010). Corporate governance, accounting and finance: A review. *Accounting & Finance,* 51(1), 96-172. Doi:10.1111/j.1467-629X.2010.00385.x

Brühl, R., & Kury, M. (2016). Rhetorical Tactics to Influence Responsibility Judgments: Account Giving in Banks Presidents' Letters During the Financial Market Crisis. *International Journal of Business Communication, 56*(3), 299-325. doi:10.1177/2329488415627356

Cerri, R., Barros, R. C., AC, P. L. F. d. C., & Jin, Y. (2016). Reduction strategies for hierarchical multi-label classification in protein function prediction. *BMC Bioinformatics, 17*(1), 373. doi:10.1186/s12859-016-1232-1

Chakraborty, V., Chiu, V., & Vasarhelyi, M. (2014). Automatic classification of accounting literature. *International Journal of Accounting Information Systems, 15*(2), 122-148. doi:10.1016/j.accinf.2014.01.001

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research, 16*, 321-357. doi:10.1613/jair.953

Chen, Y.-J., Wu, C.-H., Chen, Y.-M., Li, H.-Y., & Chen, H.-K. (2017). Enhancement of fraud detection for narratives in annual reports. *International Journal of Accounting Information Systems, 26*, 32-45. doi:10.1016/j.accinf.2017.06.004

Cho, C. H., Roberts, R. W., & Patten, D. M. (2010). The language of US corporate environmental disclosure. *Accounting, Organizations and Society, 35*(4), 431-443. doi:10.1016/j.aos.2009.10.002

Cianci, A. M., & Kaplan, S. E. (2010). The effect of CEO reputation and explanations for poor performance on investors' judgments about the company's future performance and management. *Accounting, Organizations and Society,* 35(4), 478-495. Doi:10.1016/j.aos.2009.12.002

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. doi:10.1080/00014788.2003.9729645

Clatworthy, M., & Jones, M. J. (2006). Differential patterns of textual charactheristics and company performance in the chairman's statements. *Accounting, Auditing & Accountability Journal,* 19(4), 483-511. doi:10.1108/09513570610679100

Cole, C. J., & Jones, C. L. (2004). The usefulness of MD&A disclosures in the retail industry. *Journal of Accounting, Auditing and Finance, 19*(4), 361-388. doi:10.1177/0148558X0401900401

Costa, E. P., Lorena, A. C., Carvalho, A. C. P. L. F., & Freitas, A. A. (2007). *A Review of Performance Evaluation Measures for Hierarchical Classifiers*. Paper presented at the Learning II: papers from the AAAI.

Davis, A. K., Piger, J. M., & Sedor, L. M. (2012). Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language*. *Contemporary Accounting Research, 29*(3), 845-868. doi:10.1111/j.1911-3846.2011.01130.x

Davis, A. K., & Tama-Sweet, I. (2012). Managers' Use of Language Across Alternative Disclosure Outlets: Earnings Press Releases versus MD&A*. *Contemporary Accounting Research, 29*(3), 804-837. doi:10.1111/j.1911-3846.2011.01125.x

Domingos, P., & Pazzani, M. (1997). On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. *Machine Learning, 29*, 103-130. doi:10.1023/A:1007413511361

Du, S., & Yu, K. (2020). Do Corporate Social Responsibility Reports Convey Value Relevant Information? Evidence from Report Readability and Tone. *Journal of Business Ethics*. doi:10.1007/s10551-020-04496-3

Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics, 64*(2-3), 221-245. doi:10.1016/j.jacceco.2017.07.002

El-Haj, M., Alves, P., Rayson, P., Walker, M., & Young, S. (2019b). Retrieving, classifying and analysing narrative commentary in unstructured (glossy) annual reports published as PDF files. *Accounting and Business Research, 50*(1), 6-34. doi:10.1080/00014788.2019.1609346

El-Haj, M., Rayson, P., Walker, M., Young, S., & Simaki, V. (2019a). In search of meaning: Lessons, resources and next steps for computational analysis of financial discourse. *Journal of Business Finance & Accounting, 46*(3-4), 265-306. doi:10.1111/jbfa.12378

El-Haj, M., Rayson, P., Young, S., Moore, A., Walker, M., Schleicher, T., & Athanasakou, V. (2016). Learning Tone and Attribution for Financial Text Mining. *Lrec 2016 - Tenth International Conference on Language Resources and Evaluation*, 1820-1825. Retrieved from <Go to ISI>://WOS:000526952502006

Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). *Learning from Imbalanced Data Sets*: Springer Nature Switzerland AG.

Fernandez-Delgado, M., Cernades, E., Barro, S., & Amorim, D. (2014). Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *Journal of Machine Learning Research, 15*, 3133-3181.

Fisher, I. E., Garnsey, M. R., Goel, S., & Tam, K. (2010). The Role of Text Analytics and Information Retrieval in the Accounting Domain. *Journal of Emerging Technologies in Accounting, 7*(1), 1-24. doi:10.2308/jeta.2010.7.1.1

Fisher, I. E., Garnsey, M. R., & Hughes, M. E. (2016). Natural Language Processing in Accounting, Auditing and Finance: A Synthesis of the Literature with a Roadmap for Future Research. *Intelligent Systems in Accounting, Finance and Management, 23*(3), 157-214. doi:10.1002/isaf.1386

Garzone, G. (2006). The use of discursive features expressing causal relations in annual company reports. In J. Flowerder & M. Gotti (Eds.), *Studies in specialized discourse* (pp. 81-107). Bern/New York: Peter Lang.

Goel, S., Gangolly, J., Faerman, S. R., & Uzuner, O. (2010). Can Linguistic Predictors Detect Fraudulent Financial Filings? *Journal of Emerging Technologies in Accounting, 7*(1), 25-46. doi:10.2308/jeta.2010.7.1.25

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix. *Educational Researcher, 40*(5), 223-234. doi:10.3102/0013189x11413260

Guo, L., Vargo, C. J., Pan, Z., Ding, W., & Ishwar, P. (2016). Big Social Data Analytics in Journalism and Mass Communication. *Journalism & Mass Communication Quarterly, 93*(2), 332-359. doi:10.1177/1077699016639231

Guo, L., Ma, Y., Cukic, B., & Singh, H. (2004). *Robust prediction of fault-proneness by random forests*. Paper presented at the 15th International Symposium on Software Reliability Engineering.

Henry, E. (2006). Market Reaction to Verbal Components of Earnings Press Releases: Event Study Using a Predictive Algorithm. *Journal of Emerging Technologies in Accounting, 3*(1), 1-19. doi:10.2308/jeta.2006.3.1.1

Henry, E. (2008). Are Investors Influenced By How Earnings Press Releases Are Written? *Journal of Business Communication, 45*(4), 363-407. doi:10.1177/0021943608319388

Henry, E., & Leone, A. J. (2016). Measuring Qualitative Information in Capital Markets Research: Comparison of Alternative Methodologies to Measure Disclosure Tone. *Accounting Review, 91*(1), 153-178. doi:10.2308/accr-51161

Hines R. D. (1988). Financial accounting: In communicating reality, we construct reality. *Accounting, Organizations and Society,* 13(3), 251-261. doi:10.1016/0361-3682(88)90003-7

Hooghiemstra, R. (2003). *The construction of reality: cultural differences in self-serving behaviour in accounting narratives*. Paper presented at the ERIM PhD Series Research in Management, Rotterdam.

Huang, A. H., Zang, A. Y., & Zheng, R. (2014). Evidence on the Information Content of Text in Analyst Reports. *Accounting Review, 89*(6), 2151-2180. doi:10.2308/accr-50833

Humpherys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K., & Felix, W. F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems, 50*(3), 585-594. doi:10.1016/j.dss.2010.08.009

Huysmans, J., Dejaeger, K., Mues, C., Vanthienen, J., & Baesens, B. (2011). An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models. *Decision Support Systems, 51*(1), 141-154. doi:10.1016/j.dss.2010.12.003

Ignatow, G., & Mihalcea, R. (2018). *An Introduction to Text Mining*. London: Sage Publications, Inc.

Im, T. L., San, P. W., On, C. K., & Alfred, R. (2013). *Analysing Market Sentiment in Financial News using Lexical Approach*. Paper presented at the 2013 IEEE Conference on Open Systems (ICOS), Sarawak, Malaysia.

Jayamohan, P., McKelvie, A., & Moss, T. W. (2017). Blame You, Blame Me: Exploring Attribution Differences and Impact in Family and Nonfamily Firms. *Family Business Review, 30*(3), 284-308. doi:10.1177/0894486517722887

Kimbrough, M. D., & Wang, I. Y. (2014). Are Seemingly Self-Serving Attributions in Earnings Press Releases Plausible? Empirical Evidence. *The Accounting Review, 89*(2), 635-667. doi:10.2308/accr-50628

Kiritchenko, S., Matwin, S., & Famili, F. (2004). *Hierarchical Text Categorization as a Tool of Associating Genes with Gene Ontology Codes*. Paper presented at the Fifteenth European Conference on Machine Learning, Pisa, Italy.

Kiritchenko, S., Matwin, S., & Famili, F. (2005). *Functional Annotation of Genes Using Hierarchical Text Categorization*. Paper presented at the Proceedings of the ACL Workshop on Linking Biological Literature, Ontologies and Databases: Mining Biological Semantics.

Koo, D. S., Julie Wu, J., & Yeung, P. E. (2017). Earnings Attribution and Information Transfers. *Contemporary Accounting Research, 34*(3), 1547-1579. doi:10.1111/1911-3846.12308

Koonce, L., Seybert, N., & Smith, J. (2011). Causal reasoning in financial reporting and voluntary disclosure. *Accounting, Organizations and Society,* 36(4-5), 209-225. doi:10.1016/j.aos.2011.03.006

Kosmopoulos, A., Partalas, I., Gaussier, E., Paliouras, G., & Androutsopoulos, I. (2014). Evaluation measures for hierarchical classification: a unified view and novel approaches. *Data Mining and Knowledge Discovery, 29*(3), 820-865. doi:10.1007/s10618-014-0382-x

Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence, 5*(4), 221-232. doi:10.1007/s13748-016-0094-0

Krippendorff, K. (2013). *Content analysis : an introduction to its methodology* (3rd ed.). Los Angeles ; London: SAGE.

Kruengkrai, C., Torisawa, K., Hashimoto, C., Kloetzer, J., Oh, J. H., & Tanaka, M. (2017). *Improving Event Causality Recognition with Multiple Background Knowledge Sources Using Multi-Column Convolutional Neural Networks.* Paper presented at the The 31st AAAI Conference on Artificial Intelligence, Kyoto, Japan.

Lamm, M., Chaganty, A., Jurafsky, D., Manning, C., & Liang, P. (2018). *QSRL: A Semantic Role-Labeling Schema for Quantitative Facts*. Paper presented at the Proceedings of the 1st Financial Narrative Processing Workshop.

Lawrence, A. (2013). Individual investors and financial disclosure. *Journal of Accounting and Economics, 56*(1), 130-147. doi:10.1016/j.jacceco.2013.05.001

Lehavy, R., Li, F., & Merkley, K. (2011). The Effect of Annual Report Readability on Analyst Following and the Properties of Their Earnings Forecasts. *Accounting Review, 86*(3), 1087-1115. doi:10.2308/accr.00000043

Leibbrand, M. P. (2015). The language of executive financial discourse. *Studies in Communication Sciences, 15*(1), 45-52. doi:10.1016/j.scoms.2015.03.006

Lessmann, S., Baesens, B., Mues, C., & Pietsch, S. (2008). Benchmarking Classification Models for Software Defect Prediction: A Proposed Framework and Novel Findings. *IEEE Transactions on Software Engineering, 34*(4), 485-496. doi:10.1109/tse.2008.35

Lewis, C., & Young, S. (2019). Fad or future? Automated analysis of financial text and its implications for corporate reporting. *Accounting and Business Research, 49*(5), 587-615. doi:10.1080/00014788.2019.1611730

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics, 45*(2-3), 221-247. doi:10.1016/j.jacceco.2008.02.003

Li, F. (2010a). Textual Analysis of Corporate Disclosures: A Survey of the Literature. *Journal of Accounting Literature, 29*, 143-165.

Li, F. (2010b). The Information Content of Forward-Looking Statements in Corporate Filings-A Naïve Bayesian Machine Learning Approach. *Journal of Accounting Research, 48*(5), 1049-1102. doi:10.1111/j.1475-679X.2010.00382.x

Li, F. (2010c). Managers' Self-Serving Attribution Bias and Corporate Financial Policies. doi:10.2139/ssrn.1639005

Libby, R., & Rennekamp, K. (2012). Self-Serving Attribution Bias, Overconfidence, and the Issuance of Management Forecasts. *Journal of Accounting Research, 50*(1), 197-231. doi:10.1111/j.1475-679X.2011.00430.x

Liew, W. T., Adhitya, A., & Srinivasan, R. (2014). Sustainability trends in the process industries: A text mining-based analysis. *Computers in Industry, 65*(3), 393-400. doi:10.1016/j.compind.2014.01.004

Ling, C. X., & Li, C. (1998). *Data Mining for Direct Marketing: Problems and Solutions*. Paper presented at the In Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98), New York, NY.

Liu, S.-H., Chen, S.-Y., & Li, S.-T. (2017). *Text-Mining Application on CSR Report Analytics: A Study of Petrochemical Industry*. Paper presented at the 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI).

Loterman, G., Brown, I., Martens, D., Mues, C., & Baesens, B. (2012). Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting, 28*(1), 161-170. doi:10.1016/j.ijforecast.2011.01.006

Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance, 66*(1), 35-65. doi:10.1111/j.1540-6261.2010.01625.x

Loughran, T., & McDonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research, 54*(4), 1187-1230. doi:10.1111/1475-679x.12123

Martens, D., Baesens, B., Van Gestel, T., & Vanthienen, J. (2007). Comprehensible credit scoring models using rule extraction from support vector machines. *European Journal of Operational Research, 183*(3), 1466-1476. doi:10.1016/j.ejor.2006.04.051

Melloni, G., Stacchezzini, R., & Lai, A. (2016). The tone of business model disclosure: an impression management analysis of the integrated reports. *Journal of Management & Governance, 20*(2), 295-320. doi:10.1007/s10997-015-9319-z

Merkl-Davies, D. M., & Brennan, N. M. (2007). Discretionary disclosure strategies in corporate narratives : incremental information or impression management? *Journal of Accounting Literature, 26*(116-196).

Merkl-Davies, D. M., Brennan, N. M., & McLeay, S. J. (2011). Impression management and retrospective sense-making in corporate narratives. *Accounting, Auditing & Accountability Journal, 24*(3), 315-344. doi:10.1108/09513571111124036

Moeyersoms, J., Junqué de Fortuny, E., Dejaeger, K., Baesens, B., & Martens, D. (2015). Comprehensible software fault and effort prediction: A data mining approach. *Journal of Systems and Software, 100*, 80-90. doi:10.1016/j.jss.2014.10.032

Nakano, F. K., Pinto, W. J., Pappa, G. L., & Cerri, R. (2017). *Top-down strategies for hierarchical classification of transposable elements with neural networks*. Paper presented at the 2017 International Joint Conference on Neural Networks (IJCNN).

Ogden, S., & Clarke, J. (2005). Customer disclosures, impression management and the construction of legitimacy. *Accounting, Auditing & Accountability Journal, 18*(3), 313-345. doi:10.1108/09513570510600729

Oh, J. H., Torisawa, K., Hashimoto, C., Sano, M., De Saeger, S., & Ohtake, K. (2013, 4-9 August 2013). *Why-question answering using intra- and inter- sentential causal relations.* Paper presented at the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria.

Ouyang, F. S., Guo, B. L., Ouyang, L. Z., Liu, Z. W., Lin, S. J., Meng, W., . . . Yang, S. M. (2019). Comparison between linear and nonlinear machine-learning algorithms for the classification of thyroid nodules. *Eur J Radiol, 113*, 251-257. doi:10.1016/j.ejrad.2019.02.029

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*, 2825-2830. Retrieved from <Go to ISI>://WOS:000298103200003

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., & Webber, B. (2008). *The Penn Discourse TreeBank 2.0*. Paper presented at the Proceedings of LREC 2008, 2008.

Provost, F., & Fawcett, T. (2013). *Data science for business what you need to know about data mining and data-analytic thinking*(1st ed., pp. 1 online resource (xviii, 384 pages)).

Purda, L., & Skillicorn, D. (2015). Accounting Variables, Deception, and a Bag of Words: Assessing the Tools of Fraud Detection. *Contemporary Accounting Research, 32*(3), 1193-1223. doi:10.1111/1911-3846.12089

Rennekamp, K. (2012). Processing fluency and investors' reactions to disclosure readability. *Journal of Accounting Research, 50*(5), 1319-1354. doi:https://doi.org/10.1111/j.1475-679X.2012.00460.x

Rosenkranz, J., & Pollach, I. (2016). The framing and reframing of corporate financial results. *Corporate Communications: An International Journal, 21*(1), 103-119. doi:10.1108/ccij-07-2015-0041

Săndulescu, M.-S. (2019). Self-serving performance attributions in the letters to shareholders: A longitudinal case study. *Journal of Accounting and Management Information Systems, 18*(2), 283-301. doi:10.24818/jamis.2019.02006

Scott, M. B., & Lyman, S. M. (1968). Accounts. *American Sociological Review, 33*(1), 46-62.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys, 34*(1). doi:10.1145/505282.505283

Shirata, C. Y., & Sakagami, M. (2008). An Analysis of the ''Going Concern Assumption'': Text Mining from Japanese Financial Reports. *Journal of Emerging Technologies in Accounting, 5*(1), 1-16. doi:10.2308/jeta.2008.5.1.1

Shirata, C. Y., Takeuchi, H., Ogino, S., & Watanabe, H. (2011). Extracting Key Phrases as Predictors of Corporate Bankruptcy: Empirical Analysis of Annual Reports by Text Mining. *Journal of Emerging Technologies in Accounting, 8*(1), 31-44. doi:10.2308/jeta-10182

Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond accuracy, F-Score and ROC: A family of discriminant measures for performance evaluation. *Ai 2006: Advances in Artificial Intelligence, Proceedings, 4304*, 1015-+. Retrieved from <Go to ISI>://WOS:000244891200114

Szekely, N., & Vom Brocke, J. (2017). What can we learn from corporate sustainability reporting? Deriving propositions for research and practice from over 9,500 corporate sustainability reports published between 1999 and 2015 using topic modelling technique. *PLoS One, 12*(4), e0174807. doi:10.1371/journal.pone.0174807

Tausczik, Y. R., & Pennebaker, J. W. (2009). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology, 29*(1), 24-54. doi:10.1177/0261927x09351676

Terrin, N., Schmid, C. H., Griffith, J. L., D'Agostino, R. B., & Selker, H. P. (2003). External validity of predictive models: a comparison of logistic regression, classification trees, and neural networks. *Journal of Clinical Epidemiology, 56*(8), 721-729. doi:10.1016/s0895-4356(03)00120-3

van Atteveldt, W., van der Velden, M. A. C. G., & Boukes, M. (2021). The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms. *Communication Methods and Measures, 15*(2), 121-140. doi:10.1080/19312458.2020.1869198

Van den Bogaerd, M., & Aerts, W. (2011). Applying machine learning in accounting research. *Expert Systems with Applications, 38*(10), 13414-13424. doi:10.1016/j.eswa.2011.04.172

Van Gestel, T., Suykens, J., Baesens, B., Viaene, S., Vanthienen, J., Dedene, G., . . . Vandewalle, J. (2004). Benchmarking Least Squares Support Vector Machine Classifiers. *Machine Learning, 54*, 5-32.

Vanhoeyveld, J., Martens, D., & Peeters, B. (2019). Customs fraud detection. *Pattern Analysis and Applications, 23*(3), 1457-1477. doi:10.1007/s10044-019-00852-w

Walker, M., Young, S., Athanasakou, V., El-Haj, M., Rayson, P., & Schleicher, T. (2020). Who’s the Fairest of them All?:A Comparison of Methods for Classifying Tone and Attribution in Earnings-related Management Discourse. *Working Paper*.

Weiss, S. M. (2005). *Text mining : predictive methods for analyzing unstructured information*. New York: Springer.

Yang, F., Dolar, B., & Mo, L. (2018). Textual Analysis of Corporate Annual Disclosures: A Comparison between Bankrupt and Non-Bankrupt Companies. *Journal of Emerging Technologies in Accounting, 15*(1), 45-55. doi:10.2308/jeta-52085

Zhang, S., & Aerts, W. (2015). Management's performance justification and failure to meet earnings thresholds. *Accounting and Business Research, 45*(6-7), 841-868. doi:10.1080/00014788.2015.1048771

Zhang, S., Aerts, W., & Pan, H. (2019). Causal language intensity in performance commentary and financial analyst behaviour. *Journal of Business Finance & Accounting, 46*(1-2), 3-31. doi:10.1111/jbfa.12351

Zhao, S., Liu, T., Zhao, S., Chen, Y., & Nie, J.-Y. (2016). Event causality extraction based on connectives analysis. *Neurocomputing, 173*, 1943-1950. doi:10.1016/j.neucom.2015.09.066

CHAPTER V

CONCLUSION

This dissertation has undertaken a comprehensive exploration into top management's language use in FR and SR. The dissertation is situated at the intersection of accounting, linguistics and sustainability research. The main aim of the dissertation is to investigate the use of top management's language within the context of corporate narratives with the overarching goal of enhancing the accountability and transparency of SR as a disclosure practice. Drawing upon legitimacy theory, the investigation begins with the identification of linguistic style features and rhetorical profiles in FR and SR, followed by the examination of the effect of industry affiliation and ESG controversies on IM utilization in SR, and the assessment of novel methodologies for textual analysis in corporate reporting. The dissertation delves into how the institutional environment moderates discretionary storytelling tendencies. Furthermore, the examination of visibility's role in rhetorical IM sheds light on the strategic use of SR by companies for legitimization. Finally, the dissertation explored the feasibility of machine learning methods for attributional content and framing analysis in corporate reporting. By evaluating the performance of various classifiers, the study provides a foundation for implementing machine learning methodologies for IM research, contributing valuable insights to the accounting and management literature.

1. Theoretical Contributions

This dissertation's theoretical background is based on legitimacy theory. Despite its popularity legitimacy theory has been constantly criticized for: (1) failing to address the broader political and economic context for legitimacy strategies; (2) neglecting the motives that may lead to voluntarily disclosing good SP-related content (i.e., voluntary disclosure theory); and (3) having the legitimacy, as a concept, very narrow and yet to be explored (Patten, 2019). First, although legitimacy strategies extend beyond corporate disclosures, legitimacy theory emphasizes the importance of companies aligning their performance with societal expectations, which are embedded in corporate communications through disclosures. This resonates with the findings of this dissertation: We continue to observe distinct language usage across FR and SR, as well as distinct storytelling styles and IM strategies within SR narratives. Second, as voluntary disclosure theory argues, companies are more motivated to disclose positive news to distinguish themselves from less successful competitors (Clarkson et al., 2008). On the other hand, from the legitimacy perspective, companies that poorly perform are expected to disclose more about their SP to gain acceptance (Aluchna et al., 2023; Patten, 2019). Our results show that once controversies occur, managers tend to implement IM strategies for justification and sense-giving reasons. Consequently, the findings of this dissertation further provide empirical evidence legitimacy theory is still prominent despite the persistent critique. Third, Patten (2019) further argues that there may be value in exploring how SR may serve differently for diverse stakeholders through different organizational façades. This dissertation provides important insights into the concept of legitimacy and legitimacy strategies by analyzing top management's language use in both FR and SR, which aim at different stakeholders. Specifically, this dissertation contributes to the literature by highlighting that employing the generic term "legitimacy" without discerning these distinctions may lead to neglect of essential nuances linked to linguistic variations in corporate reporting due to diverse influences. Our contribution to the literature lies in highlighting the potential benefits of distinguishing stakeholder motives based on legitimacy types, specifically moral versus pragmatic legitimacy.

Theoretically, this dissertation also adds to the IM literature. Our studies show that we observe three main rhetorical streams in corporate narratives. The findings regarding both informative (i.e., rational-factual and analytic appeals) and storytelling (i.e., personalized sense-giving, assertive, defensive) appeals provide empirical evidence that SR is definitely a hybrid disclosure, in which both informative and promotional functions are utilized. The utilization of these strategies (e.g., defensive as a more explanatory sense-giving purposes) provides further empirical evidence that IM serves both self-presentational and information-sharing purposes, in line with prior research (Yan et al., 2019). Consequently, this dissertation highlights that IM practices in corporate narratives are not supposed to be perceived as unethical strategies. However, we acknowledge that IM can be considered unethical when deception occurs (Provis, 2010). In order to comprehend the ethical aspects of IM in SR, this dissertation may be insufficient, as the ethical evaluation can be more complex and may, for example, require assessment of SP aspects through a checklist approach. Furthermore, our research adds to the literature highlighting that thematic content and linguistic style are interlinked. In applied linguistics literature, it is argued that linguistic style (also referred to as metadiscourse) is not independent but is also related to the content of the text (Hyland, 1998). The significance of thematic differences is anticipated to be substantial and should be considered when analyzing language use in corporate reporting, in line with prior research, which shows different selective disclosure patterns within specific sustainability pillars (Roszkowska-Menkes et al., 2024). This dissertation, therefore, also highlights the importance and significance of controlling for thematic content differences in corporate narratives.

2. Practical Contributions

The empirical findings in this dissertation provide a new understanding of language use in FR and SR. Practically, our findings are expected to contribute to enhancing the quality of SR guidelines, such as GRI or EU CSRD directives, efficiently narrowing potential spaces for information asymmetry in narrative sections. The regulations are developed to invest in these tailored communication strategies of companies. While CSRD extends the requirements of SR disclosure to SMEs[41] and emphasizes the principle of double materiality, whereby sustainability issues must be significant to the company, narrative sections persist as one of the most discretionary components. Our emphasis underscores the imperative pursuit and development of transparency and accountability in the domain of SR. In addition, recognizing the genre differences, and influence of industry affiliation and ESG controversies on IM can guide policymakers in evaluating corporate communication strategies. As our results show, organizations, particularly those operating in industries with lower scrutiny, can strategically employ IM, which should be additionally considered for transparency and accountability. The findings in this dissertation show the necessity for regulatory bodies to update SR guidelines to consider regional institutional contexts and industry-specific factors, thereby enhancing transparency and mitigating discretionary disclosure tendencies in SR narratives. We highlight, furthermore, that stakeholders may consider critically evaluating SR narratives, to discern between opportunistic storytelling and genuine commitment to sustainability. Finally, we benchmark machine learning algorithms for attributional content and framing analysis, which presents a practical avenue for improving the efficiency of narrative disclosure analysis. Machine learning methods can offer auditors and investors and effective means to assess corporate communication; and importantly, we highlight that such methodologies are not limited to practical application but extend to academic literature providing a valuable tool for scholarly research and analysis. Overall, our results are expected to aid socially responsible investors and auditors in comprehensively assessing corporate narratives, including the verbal IM aspects, i.e., rhetorical, thematic and attributional, and aid investment analysts in identifying potential red flags about corporate communication.

[41] The proposal expands the qualifying criteria for SR to encompass companies with a workforce of 250 or more employees and/or net turnover exceeding 40 million Euros and/or a balance sheet surpassing 20 million Euros (CSRD - 2022/2464/EU)

3. Limitations and future recommendations

A major limitation of the studies within this dissertation relates to the representativeness of the sample which predominantly consists of FR and SR from large organizations from the EU, UK, and USA. This limitation may raise concerns about the adequacy of the sample in capturing the nuanced diversity of corporate communication practices globally. Consequently, the prevailing focus on large companies and specific geographical regions may restrict the generalizability of findings to a broader spectrum of corporate communication strategies globally. Furthermore, the utilization of machine learning methods, particularly supervised algorithms, may introduce a challenge of inherent bias, as they exhibit bias based on the characteristics of the training data. This may pose a challenge to the applicability of algorithms between corporate reporting genres, such as training data received from FR, but classification used on SR narratives. Further research may aim to ascertain the adaptability and transferability of machine learning methodologies across diverse corporate reporting genres.

Another significant limitation within this dissertation stems from the paucity of controllable individual-level CEO characteristics, a circumstance primarily due to the insufficient availability of data. In the literature, the primary focus of the literature centers on portraying IM as a strategic decision at the firm-level as a response to external motives such as stakeholder scrutiny or institutional environment, representing a deliberate and conscious managerial implementation, with limited attention to the underlying individual-level motivations for voluntary CSR disclosure behaviors (Bolino et al., 2016; Lassoued & Khanchel, 2022). We recognize that CEO-signed letters may not always reflect the CEO's writing and involve input from public relations specialists, yet CEOs heavily influence the content and language style (Post et al., 2022). Studies show that changes in management (e.g., CEO) lead to changes in the style, length and content of these letters (Eggers & Kaplan, 2009).

Consequently, both in management and sustainability research one of the most popular discussions is whether and how much individual-level factors (e.g., CEO personality) influence corporate actions and outcomes, including sustainability communication (Wernicke et al., 2022; Ruiz-Blanco et al., 2022). This holds particular significance within the context of SR due to its discretionary nature as SR narratives contain personalized messages and business tales from top management, addressing key corporate events, achievements and future prospects as presented by the corporate leader (Amernic & Craig, 2007; Fuoli, 2018). Personalized language in corporate communications reflects individual traits such as thinking style or emotional status, which may affect the use of IM in SR narratives despite the effect of firm-level strategic and legitimacy needs (Hyland, 1998; Mahmoudian et al., 2021; Venugopal et al., 2023). Theoretically, upper echelons scholars argue that leaders' background characteristics and personalities significantly shape their interactions, motivation and influence on stakeholders and CSR practices of firms (Chin et al., 2013; Prömpeler et al., 2023). Further research may investigate how individual-level factors influence IM and disclosure practices within SR, which may provide a nuanced view of how CEOs take sustainability initiatives, translate them into strategic communications within SR, and ultimately impact stakeholder perceptions and decisions. Given that individual characteristics, such as personality traits, are also reflected in the language (Hyland, 1998), we posit that the use of linguistic style, and consequently rhetorical profiles, is not solely a firm-level strategy but may also encompass elements of individual-level motives. For example, a strong narcissism in a CEO's personality may lead to a higher use of assertive and powerful style, and a strong conscientiousness may lead to a higher use of defensive and explanatory style. This comprehension would not only add depth to the understanding of corporate sustainability communications but also hold implications for transparency, stakeholder management and the credibility of sustainability efforts.

Moreover, further research may look into how IM strategies affect SRI decisions, which represents a more focused stakeholder group and prominent target audience for SR. As SRI value the financial stability and financial materiality of corporate CSR strategies (Jansson & Biel, 2014; Puriola & Mäkelä, 2019), the communication strategies related to CSR strategies and financial (im)materiality could potentially open a new research avenue. The value of financial materiality given by investors also relates to pragmatic relationship between SRI and organization's SP. Pragmatic legitimacy demanded by SRI may prompt managers to communicate regarding financial materiality of sustainability-related information in a manner that emphasizes transparency and accountability. Finally, in light of previous research highlighting the shortcomings in current SR initiatives, such as GRI, regarding transparency and accountability (Roszkowska-Menkes et al., 2024), conducting a comparative investigation into the impact of EU CSRD regulations on the utilization of IM may represent a compelling research area.

Reference List

Aluchna, M., Roszkowska-Menkes, M., Jastrzebska, E., & Bohdanowicz, L. (2023). Sustainability reporting as a social construct: the systematic literature review within socio-political view. *Social Responsibility Journal, 19*(8), 1535-1554. https://doi.org/10.1108/SRJ-06-2022-0231

Amernic, J. H., & Craig, R. J. (2007). Guidelines for CEO-speak: editing the language of corporate leadership. *Strategy & Leadership, 35*(3), 25-31. https://doi.org/10.1108/10878570710745802

Bolino, M., Long, D., & Turnley, W. (2016). Impression Management in Organizations: Critical Questions, Answers, and Areas for Future Research. *Annual Review of Organizational Psychology and Organizational Behavior, 3*(1), 377-406. https://doi.org/10.1146/annurev-orgpsych-041015-062337

Chin, M. K., Hambrick, D. C., & Treviño, L. K. (2013). Political Ideologies of CEOs. *Administrative Science Quarterly, 58*(2), 197-232. https://doi.org/10.1177/0001839213486984

Clarkson, P. M., Li, Y., Richardson, G. D., & Vasvari, F. P. (2008). Revisiting the relation between environmental performance and environmental disclosure: An empirical analysis. *Accounting, Organizations and Society, 33*(4-5), 303-327. https://doi.org/10.1016/j.aos.2007.05.003

Directive (EU) 2022/2464 of the European Parliament and of the Council of 14 December 2022 amending Regulation (EU) No 537/2014, Directive 2004/109/EC, Directive 2006/43/EC and Directive 2013/34/EU, as regards corporate sustainability reporting. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022L2464.

Eggers, J. P. & Kaplan, S. (2009). Cognition and renewal: Comparing CEO and organitazional effects on incumbent adaptation to technical change. *Organization Science, 20*(2), 461-477. https://doi.org/orsc.1080.0401

Fuoli, M. (2018). Building a Trustworthy Corporate Identity: A Corpus-Based Analysis of Stance in Annual and Corporate Social Responsibility Reports. *Applied Linguistics, 39*(6), 846-885. https://doi.org/10.1093/applin/amw058

Hyland, K. (1998). Exploring Corporate Rhetoric: Metadiscourse in the CEO's Letter. *The Journal of Business Communication, 35*(2), 224245. https://doi.org/10.1177/002194369803500203

Jansson, M., & Biel, A. (2014). Investment Institutions' Beliefs About and Attitudes Toward Socially Responsible Investment (SRI): A Comparison Between SRI and Non-SRI Management. *Sustainable Development, 22*(1), 33-41. https://doi.org/10.1002/sd.523

Lassoued, N., & Khanchel, I. (2022). Voluntary CSR disclosure and CEO narcissism: the moderating role of CEO duality and board gender diversity. *Review of Managerial Science, 17*(3), 1075-1123. https://doi.org/10.1007/s11846-022-00555-3

Mahmoudian, F., Nazari, J. A., Gordon, I. M., & Hrazdil, K. (2021). CEO personality and language use in CSR reporting. *Business Ethics, the Environment & Responsibility, 30*(3), 338-359. https://doi.org/10.1111/beer.12347

Patten, D. M. (2019). Seeking legitimacy. *Sustainability Accounting, Management and Policy Journal, 11*(6), 1009-1021. https://doi.org/10.1108/sampj-12-2018-0332

Post, C., Lokshin, B., & Boone, C. (2022). What changes after women enter top management teams? A gender-based model of strategic renewal. *Academy of Management Journal, 65*(1), 273-303. https://doi.org/10.5465/amj.2018.1039

Provis, C. (2010). The ethics of impression management. *Business Ethics: A European Review, 19*(2), 199-212. https://doi.org/10.1111/j.1467-8608.2010.01584.x

Prömpeler, J., Veltrop, D. B., Stoker, J. I., & Rink, F. A. (2023). Striving for sustainable development at the top: Exploring the interplay of director and CEO values on environmental sustainability focus. *Business Strategy and the Environment*. https://doi.org/10.1002/bse.3408

Puroila, J., & Mäkelä, H. (2019). Matter of opinion: Exploring the socio-political nature of materiality in disclosures in sustainability reporting. *Accounting, Auditing & Accountability Journal, 32*(4), 1043-1072. https://doi.org/10.1108/aaaj-11-2016-2788

Roszkowska-Menkes, M., Aluchna, M., & Kaminski, B. (2024). True transparency or mere decoupling? The study of selective discloure in sustainability reporting. *Critical Perspectives on Accounting, 98,* 1-24. https://doi.org/10.1016/j.cpa.2023.102700

Ruiz-Blanco, S., Romero, S., & Fernandez-Feijoo, B. (2022). Green, blue or black, but washing-What company characteristics determine greenwashing? *Environment, Development and Sustainability, 24*, 4024-4045. https://doi.org/10.1007/s10668-021-01602-x

Venugopal, A., Nerur, S., Yasar, M., & Rasheed, A. (2023). CEO personality and corporate sustainability performance. *Management Decision, Ahead-of-print*(ahead-of-print). https://doi.org/10.1108/MD-06-2022-0842

Wernicke, G., Sajko, M., & Boone, C. (2022). How much influence do CEOs have on company actions and outcomes? The Example of corporate social responsibility. *Academy of Management Discoveries, 8*(1). https://doi.org/10.5465/amd.2019.0074

Yan, B., Aerts, W., & Thewissen, J. (2019). The informativeness of impression management − financial analysts and rhetorical style of CEO letters. *Pacific Accounting Review, 31*(3), 462-496. https://doi.org/10.1108/par-09-2017-0063