Determinants of Narrative Content of Sustainability Reporting by Japanese Companies

Yuriko Nakao

Corporate Communication Division, Cetus & General Press Inc.

and Department of Economics, Kwansei Gakuin University, Japan

Kimitaka Nishitani

(Research Institute for Economics and Business Administration, Kobe University, Japan

Katsuhiko Kokubu

Graduate School of Business Administration, Kobe University, Japan

Determinants of Narrative Content of Sustainability Reporting by Japanese Companies

Abstract

In this study, we explore environmental performance and other determinants of sustainability reporting by Japanese companies using an estimation method that accounts for sample selection bias. We applied the computer-aided text analysis software DICTION to a narrative content of those reports with a particular focus on two linguistic themes: "Optimism" and "Certainty." These analyses yielded several notable results. First, we found that companies with high environmental and social performance typically engage in sustainability reporting. Second, we showed that the narrative content of reports released by companies with low environmental performance tend to contain expressions related to "Optimism," whereas reports released by companies with high environmental performance tend to use expressions related to "Certainty." Finally, we found no effect of social performance on the narrative content of company reports.

1. Introduction

Although legislative regulations related to corporate social and environmental disclosure do not exist in Japan, companies' growing involvement in environmental issues precipitated an increase in environmental reporting in the second half of the 1990s. This trend continued into the early 2000s, as a number of Japanese companies began to issue sustainability reports that included information related to socially-relevant issues.[1] A survey conducted by KPMG (2011) indicated that 69% of the top 100 grossing companies in 34 countries engage in sustainability reporting. Of all nations surveyed, Japan had the second-highest proportion (99%) of companies that report information related to environmental and social issues. This suggests that Japan is relatively advanced in terms of sustainability reporting.

Although guidelines exist for sustainability reporting, it is voluntary, and thus is left to the company's discretion. The contents of sustainability reports, as well as the stakeholders targeted as preferential readers, vary according to the issues addressed in the report. Whereas environmental reporting may describe policies designed to curb global warming or manage waste material, social reporting may describe actions related to employees or trade partners. However, mere quantitative data related to environmental and/or social issues are often difficult to interpret. As such, qualitative information in the sustainability reports assumes a greater importance than the data contained in financial reports. Despite its significance, the content of sustainability reports has yet to be comprehensively investigated. The nature of the information contained in sustainability reports, as well as its determinants, thus warrant empirical investigation. Given this, the objective of this paper is to elucidate the determinants of sustainability reporting by Japanese companies, as well as to describe the information contained in the reports' narrative content.

The remainder of the paper is organized as follows. Section 2 describes extant research related to sustainability reporting and text analysis and presents the objectives of the current study. Section 3 explains the scoring process we employed to analyze the narrative content of sustainability reports. Section 4 describes the variables used in the current study and the methodology we employ. Section 5 discusses the analysis results, and Section 6 concludes the paper.

2. Review of previous research and objectives of the analysis

There is a significant amount of research on the determinants of information disclosure in Japanese sustainability reports (Kokubu et al., 2002; Higashida et al., 2005; Kokubu et al., 2012). These studies have shown that the degree to which a company's stakeholders influence business decisions affects the quality of that company's sustainability reports. In these studies, however, "quality" is determined only by the presence or absence of disclosed items.

In contrast, research concerning narrative text has featured not only sustainability reports, but also documents related to financial reporting (e.g., annual reports). Like the work on sustainability

3

reports, the research conducted on the narrative content of annual reports has also been substantial. Neu et al. (1998) attempted to determine the influence of external pressure on environmental disclosures included in annual reports of Canadian companies in environmentally sensitive industries. Based on their results, Neu et al. (1998) found that external pressures, such as regulators and environmentalists, were associated with environmental disclosures. Additionally, they suggested that it was possible to manage public impressions with the textually mediated environmental disclosures contained in annual reports. Merkl-Davies and Brennan (2007) reviewed the narrative content of information disclosed by companies in broad terms and considered narrative content to be discretionary disclosures. The authors did this to mitigate the asymmetry inherent to the information disclosed to the company's external stakeholders. They found that to control external perceptions (i.e., reputation, environment, social performance) of the company when a company experiences negative outcomes, additional discretionary information is manipulated.[2]

The advent and growing popularity of analysis software in Japan has fostered research on computer-based text analysis of financial reports (Kida, 2006; Shirata, 2008). Oda and Mihashi (2010), for example, used text mining methods for grouping. They found that specific management concepts produce a positive effect. As in this research, Oda and Mihashi also considered non-financial information in their analyses. In addition, Kitora and Okuda (2009) used text mining methods to analyze the content of CSR policies to classify companies into one of four categories, depending on the company's governance type. Shino (2010) focused on labor-related topics that are often included in sustainability reports to categorize the language in terms of three themes: "for what," "to whom," and "what to do." Because there is significant ambiguity related to these terms in the Japanese language, text analysis were employed to accurately classify the text. Although these studies have proven useful, some have exclusively focused on text related to specific managerial concepts, CSR policies, or CSR-related labor, and not entire sustainability reports.

Rather than focus on specific components of financial reports, Cho et al. (2010) analyzed the textual information (in its entirety) of American companies' 10-K annual reports. Cho et al. (2010, p.434) developed two hypotheses to demonstrate that the tone and context related to the environment disclosure is correlated to the quality of company's environmental performance evaluation. Utilizing the DICTION software to determine "Optimism" and "Certainty" for the environment disclosure, their hypotheses as follows,

H1: The "Optimism" exhibited in 10-K report environmental disclosures will be negatively related to firm environmental performance.

H 2: The "Certainty" exhibited in 10-K report environmental disclosures will be positively related to firm environmental performance.

They supported both hypotheses that companies that perform poorly with respect to environmental CSR often use ambiguous expressions instead of clear language. To gauge environmental

performance Cho et al. (2010) used the environmental rating measure used by KLDas of 2002.[3]

In this paper, we utilize entire sustainability reports issued by Japanese companies as sources for narrative data. As such, we follow the methods of Cho et al. (2010), which used the narrative content of environmental disclosures. Doing so will allow us to clarify whether Cho et al.'s results can be replicated with Japanese companies. Although we replicate the methods used by Cho et al. (2010), there is a fundamental difference implicit to the current research. Cho et al. (2010) focused on environmental disclosures in the context of American annual reports (which are intrinsically financial in nature). As a result, the target readers of their narrative data were investors interested in the company. Contrarily, sustainability reports used as data for this study are of interest to multiple stakeholders: investors, employees, customers, trade partners, NGO's, civil communities, and the government. Moreover, while Cho et al. (2010) focused solely on environmental information, we address other reports that have been characterized as important for sustainability reporting. Given these differences, we expect our results to deviate from those of Cho et al. (2010).

We might expect that "Optimism" language employed in a company's sustainability reports is masking the company's true environmental and social performance, particularly if the company is performing poorly in either or both of these areas. We state hypothesis H1 as follows:

H1. The "Optimism" language exhibited in sustainability report will be negatively related to firm environmental and social performance.

We might expect that "Certainty" language employed in a company's sustainability reports reflects a positive environmental and social performance. We state hypothesis H2 as follows:

H2. The "Certainty" language exhibited in sustainability report will be positively related to firm environmental and social performance.

3. Sustainability reporting and scoring based on narrative content

We define sustainability reports as those reports released by a company that are related to environmental and social welfare. For the purposes of this paper, a report can also be deemed a "sustainability report" if it satisfies any of following three conditions: 1) an "environmental report"[4] that addresses only environmental issues; 2) the report that is published on   HTML or PDF file; 3) integrated report that encompasses sustainability reports and annual reports.[5] Our analyses utilize sustainability reports that have been published in English in 2011.

We focus on environmental disclosures and social disclosures in sustainability reports as narrative contents. However, to isolate narrative content as a means to facilitate narrative analysis, those sections of environmental and social disclosures containing data such as figures and tables are excluded. Further, we did not analyze narrative data related to a company's social contribution activities in the Sustainability Report. In addition, the inclusion of social contribution activities would generate a dataset too large to feasibly analyze. We excluded social contribution activities

from analysis because social activities carried out by Japanese companies are not included in the GRI (2006), an international initiative related to information disclosure.[67]

The operational definitions and boundaries described above provide a basis for evaluating narrative content using quantitative metrics. Narrative content was scored using the computer-aided text analysis software DICTION, which has been widely-adopted for research that explores narrative content in the areas of accounting and management (e.g., Sydserff and Weetman, 2002; Jeremy and Timothy, 2007).

DICTION is a computer-aided text analysis software package that assigns scores to verbal tones to evaluate the effects of word choice. DICTION simultaneously analyzes narrative content through the use of five primary semantic variables, each of which contains 35 sub-variables. It can process up to 1,500,000 words and 3,000 passages (Hart and Carroll, 2010).

The amount of narrative content in a sustainability report is left to the issuing company's discretion, which can lead to reporting disparities among different companies. In the current paper, we scored narrative content on the basis of "averaged 500-word units" to normalize variations in the amount of narrative content contained in the reports.[8, 9]

4. Research/Design

4.1 Analysis method

Our analyses were similar to those of Cho et al. (2010), with a few notable differences. First, in addition to analyzing the narrative content contained in the sustainability reports, we also focus on whether relevant information is actually disclosed. If we consider that only companies with high environmental or economic performance publish sustainability reports, determining the predictors of narrative content based exclusively on samples taken from such companies would result in sample selection bias. As such, the effects of narrative predictors could be overestimated.

Given this, sample selection bias was corrected by using Heckman's two-step method (i.e., Heckit). In the first step of Heckit, we performed a binary selection model estimation related to the existence of sustainability reporting by Japanese companies. In the second step, after adding the Inverse Mills ratio (the ratio of the probability density function to the cumulative distribution function) obtained from the first step, we estimated determinants of narrative content in environmental and social disclosures. However, when sample selection bias was not an issue, the probit model and ordinary least squares (OLS) method were independently used for the first and second steps, respectively.

The second deviation from Cho et al. (2010) relates to the assumed determinants of environmental (and social) disclosures' narrative content. In Cho et al., the authors showed that environmental performance influences the narrative content of environmental disclosures. In this study, we consider that in addition to environmental (and social) performance, other factors affect the nature of the narrative content in environmental (and social) disclosures. As such, we include a number of factors

from prior research that have been demonstrated to affect the quality of environmental information. In this way, we provide an exploratory perspective on what factors may regulate the narrative content of Japanese environmental and social disclosures.

Finally, we add explanatory variables that have been used in prior research focused on Japanese companies. Cho et al. (2010) did not do this.

4.2 Range of sample companies

Following Cho et al. (2010), we adopt the SRI survey rating. To measure environmental, social, and governmental (ESG) performance, we use STOXX's[10] leader index.[11] STOXX's ESG ratings were determined on the basis of survey data obtained by Sustainalytics,[12] a European SRI research agency. In the first step of Heckit, 260 out of 315 Japanese companies listed in STOXX's ESG listings served as analysis subjects. Note that we excluded those companies that practiced different forms of accounting such as financial and insurance sector, and those for which there were missing values. In the second step of Heckit, 164 companies that (a) were listed in STOXX's ESG ratings, and (b) publish sustainability reports in English served as analysis subjects.[13] However, 11 companies do not disclose the social dimension in sustainability reports. Therefore 153 companies disclose only the environmental dimension.

4.3 Potential determinants of sustainability reporting

In the binary choice model we used in the first step of the Heckit estimation, the outcome measure was the existence of sustainability reporting (1 if yes, 0 if no). Regression using Heckit estimation requires the incorporation of predictor variables that influence the existence of sustainability reporting, but do not directly affect the narrative content itself. We identified two predictor variables that meet these requirements.

First, because the production of sustainability reports is voluntary on the part of the issuing company, we expect the promotion of environmental and social activities influence the presence or absence of sustainability reports. The United Nations Global Compact (UNGC) is an international program related to the managerial practice of voluntary environmental and social activities. On the basis of ten principles, these activities fall into four categories: human rights, labor, environment, and anti-corruption. To date, approximately 7,000 companies in 145 countries are members of the UNGC. Since UNGC membership is expected to affect whether a company releases sustainability reports, a dummy variable was defined indicating membership in the UNGC or lack thereof. UNGC member companies were assigned a value 1; non-members were assigned a value 0. Member companies of the UNGC are listed on the UNGC's website. We used this list to determine a company's membership in the Global Compact.

Second, sustainability reports contain information that spans several departments and relates to a

variety of CSR-related activities. Therefore, it is important to consider (a) whether a company possesses a corporate system and (b) the amount of time which the system has been in place. To explore whether "corporate age" affects the likelihood of a firm to produce a sustainability report, we also utilize the logarithm of corporate age as a predictor variable.

4.4 Variables related to narrative content in environmental and social disclosures

In the second step of the Heckit regression analysis, we utilize the computer-aided text analysis software, DICTION, to determine "Optimism" and "Certainty" as dependent variables for representative, narrative content contained within sustainability reports. DICTION defines "Optimism" as the extent to which a text reflects praise, satisfaction, and inspiration rather than blame, hardship and denial. DICTION measures "Certainty" as the extent to which a text is tenacious, leveling, insistent, based on collectives or numerical terms rather than ambivalent, and filled with self-references or variety.

Short and Palmer (2008, p. 732), who reviewed research that used DICTION in the field of management strategy, found that the semantic variable "Optimism" relates to the "CEO's excessive self-confidence and arrogance". Cho et al. (2010, p. 434) argued that in American 10-K environmental disclosures, "Optimism" correlates negatively with the company's environmental performance.

In this paper, we treat "Optimism" as a variable that represents narrative content that communicates hopefulness and ambiguity. By employing hopeful and/or ambiguous language in its disclosures, a company can mask the company's true environmental and social performance, particularly if the company is performing poorly in either or both of these areas.

According to Demers and Vega (2011, p.35), who analyzed the effect of language tone in periodically-published stock price results, past optimistic predictions were sometimes negated through the use of language reflecting "Certainty," the second principal semantic variable. Cho et al. (2010, p. 34) discovered that the appearance of language reflecting "Certainty" is positively related with a company's environmental performance. Given these findings, we consider "Certainty" as an indicator of positive expression in the narrative content. As such, these variables serve as proxy indicators for environmental and social performance.

4.5 Factors thought to affect sustainability reporting and their narrative content

Following the methods of Cho et al. (2010), we use scores related to STOXX's environmental and social performance ratings as proxy measures for environmental and social performance to explore the possibility that environmental and social performance affects the narrative content that is disclosed in sustainability reports.

Given that Mallin et al. (2012) found that corporate governance is positively correlated with

environmental and social information disclosure, we use STOXX's governance rating as a proxy variable for governance performance.

Similarly, in the tradition of research that has examined the association between sentence to page number ratios and the quality of environmental disclosures (see Hooks and Staden, 2011), we adopted the number of pages in environmental and social disclosures to represent those reports' respective narrative content.

A significant amount of research from other countries (e.g., Leszczynska, 2012; Skouloudis, 2012) has attempted to compare disclosure ratios among companies on the basis of references provided by the Global Reporting Initiative's Sustainability Reporting Guidelines (GRI, 2006). As approximately 2000 companies refer to the GRI guidelines, we explore whether it affects the narrative content in a company's environmental and social reports. Specifically, we use a dummy variable that takes the value 1 if the GRI guideline is consulted and 0 if it is not.[14] We believe that adherence to the GRI guideline should only influence the narrative content and not management's decision to engage in sustainability reporting.

Japanese scholars have conducted a significant amount of research related to determinants of information disclosure in sustainability reports (Kokubu et al, 2002; Higashida et al, 2005; Kokubu et al., 2012). On the basis of this research, we assume that stakeholders influence the narrative content of sustainability reports. Given this, we consider several other variables that may affect the content of sustainability reports. These variables include: the ratio of advertisement expenses to net sales (an indicator of consumer relevance), the free float weight (a measure of dependence on the capital market), the ratio of shareholding by financial institutions (which indicates the influence of banks), and the ratio of shareholding by foreign investors (which measures the influence of foreign investors).

Cho et al. (2010) showed that in the heavy manufacturing industry, which causes significant environmental damage during production, narrative content related to the environment is particularly affected. As a result, we use a dummy variable indicating whether the company is in the manufacturing industry.

In addition to the predictor variables described above, we also employed variables to control for the possibility that environmental and social performance is correlated with narrative content in sustainability reports. Following Cho et al. (2010), the logarithm of total capital is used as a proxy variable for the size of the firm, the return on assets (ROA, which is net income divided by total assets) is used as a proxy for profitability, and leverage (debt divided by total assets) is used as a proxy for corporate safety.

We retrieved financial information for Japanese companies from the Nikkei NEEDS – Financial Quest database. We targeted sustainability reports that were released in 2011, and the contents of the environmental and social disclosures were from 2010. Energy consumption related to environmental

activities and the volume of global warming emissions are related to net sales in a given year. Therefore, we use financial data from 2010 – the year in which the environmental and social activities were performed. We use these variables in the verification analysis. We set up the model as follows :

$$EDC_i = x_i^{'}\beta + u_i \qquad (1)$$

$$x_i^{'}\alpha + v_i^{'}\gamma + \varepsilon_i > 0 \qquad (2)$$

where $u \sim N(0,\sigma)$, $\varepsilon \sim N(0,1)$, $corr(u,\varepsilon)=\rho$, x : explanatory variables except instrumental variables, v : instrumental variables, and α, β and γ : parameters estimated.

Eq. 1 is performance model and Eq. 2 is selection model. Heckit estimate Eqs. 1 and 2 simultaneously, in which the inverse Mills ratio calculated in Eq. 1 is included in Eq. 2. When Eqs. 1 and 2 are estimated independently, $v_i^{'}\gamma$ is not included in Eq. 2.

5. Estimation results

Table 1 summarizes the descriptive statistics for all the variables included in the models. According to Table 1, there is almost no difference between average values of Optimism and Certainty in environmental and social disclosures. Moreover, the standard deviations for Optimism and Certainty are not large. In terms of predictor variables, the average environmental rating was 58.9, and the social rating that was 41.5 and the governance rating that was 44.8 were comparatively lower. In addition, the environmental disclosures were 13.2 pages long, on average; social disclosures averaged 10.4 pages in length.

(Table 1: Descriptive statistical data)

Table 2 outlines the results of the model estimation. The first stage (i.e., the selective model) shows determinants for the existence of sustainability reporting; the second stage (i.e., the performance model) presents estimated results for the determinants of sustainability reports' narrative content. Proxy variables used to represent narrative content are: "Optimism" in environmental disclosures for model (1), "Certainty" in environmental disclosures for model (2), "Optimism" in social disclosures for model (3), and "Certainty" in social disclosures for model (4). For these models, the Heckit estimation results are shown when a sample selection bias is identified through the use of a Wald test. When no bias is identified, probit model estimation results are shown for the first stage and OLS estimation results are shown for the second stage.

(Table 2: Estimation results)

Let us first turn to the results from the first stage of analysis, which was related to the issuance of sustainability reports. In models (1) and (2), environmental rating and membership in the UNGC were significantly and positively related to the issuance of sustainability reports (p < .01). The size of the firm was a marginally significant predictor (p < .10). These results indicate that companies that (a) perform well with respect to environmental activities, (b) are large, or (c) participate in international initiatives related to environmental and social activities tend to engage in sustainability reporting. Moreover, in models (3) and (4), social rating, size of the firm, membership in the UNGC, and whether the firm is geared towards manufacturing are positive predictors of sustainability reporting (p < .01). These results provide evidence that companies that (a) engage in socially-responsible activities, (b) are large in size, (c) participate in international initiatives related to environmental and social activities, or (d) belong to the manufacturing industry tend to engage in sustainability reporting. Because the dummy variable for the manufacturing industry is significantly positive only for models (3) and (4), it seems as though manufacturing firms engage in sustainability reporting that includes both environmental and social information.

Next, let us turn the estimation results for the performance model. For those environmental disclosures in which "Optimism" was treated as the outcome variable, environmental rating and governance rating were shown to be significant negative predictors (p < .01). Leverage was similarly shown to be a significant negative predictor, albeit to a slightly weaker degree (p < .05). In contrast, the dummy GRI variable was a significant positive predictor (p < .05). These results indicate that companies with low environmental performance, weak governance, high safety, or consult the GRI guideline tend to use language related to "Optimism" in their environmental disclosures. For those environmental disclosures in which "Certainty" was treated as the dependent variable, environmental rating was a marginally-significant positive predictor (p < .10). Governance rating was similarly shown to be a marginally-significant predictor of language related to "Certainty," but the association was negative (p < .10). Stated simply, companies with good environmental performance and low governance tend to include expressions related to "Certainty" in the content of their environmental disclosures.

In our analysis of model (3), we found that the ratio of shareholding by financial institutions was significantly and positively related to the presence of language related to "Optimism" in social disclosures (p < .10). This suggests that companies that are characterized by greater influence of financial institutions tend to include terms related to "Optimism" in the sections of their sustainability reports related to social responsibility. We discovered no other significant associations with language related to "Optimism." Similarly, we found no evidence for relationships between the

predictor variables and language related to "Certainty" in social disclosures. As a result of these limited findings, it remains difficult to explain or interpret the inclusion of certain social information in sustainability reports.

The coefficients associated with R^2 are low in performance models (2) to (4). This may be a result of the relatively small variances associated with the outcome variables (as illustrated in Table 1).

6. Conclusion

In this paper, we explored the determinants of (a) sustainability reporting by Japanese companies, and (b) the narrative content of those sustainability reports. Unlike previous research, however, we used an estimation method that accounts for sample selection bias. The main findings are as follows.

First, we showed that companies that perform well on environmental or social indicators, join the UNGC, and/or are large in size tend to issue sustainability reports. These results reinforce the findings of Kokubu et al. (2002) that large companies are more prone to social pressure. In addition to reinforcing past findings, however, we can also conclude that membership in the UNGC program also influences the decision to issue a sustainability report. This represents a novel finding for research involving Japanese companies.

Second, we found that companies increase the degree to which they use ambiguous and/or optimistic expressions in their sustainability reports when environmental performance is low. This finding replicates those of Cho et al. (2010).

Third, we found that governance performance negatively influences the degree to which companies use language related to "Optimism" and "Certainty" in their environmental disclosures. This indicates that companies that perform poorly with respect to governance tend to use expressions containing ambiguity and optimism in their environmental disclosures. Curiously, the same negative relationship exists for corporate governance and "Certainty," which serves as an indicator for language that is direct and unambiguous. These inconsistencies demand continued research in this area.

Fourth, we found no significant association between social performance and narrative content (written in social disclosures) or environmental performance and narrative content (in environmental disclosures). This may be explained by the fact that Japanese stakeholders are not terribly interested in the content of social disclosures. As such, companies neglect to use the expected descriptive expressions.

Similarly, and in contrast with previous research on Japanese companies, we found no significant influence of the structural makeup of company stakeholders on the narrative content of environmental and/or social disclosures. This result may be explained by the fact that the variables related to narrative content are restricted to "Optimism" and "Certainty." The inclusion of other semantic indicators may reveal significant associations with corporate stakeholder structure. To

further elucidate the association between stakeholders and narrative content, future researchers should consider changing the definitional features of the outcome expressions they employ.

Our analyses concerning the relationship between governance rating and the narrative content in Japanese environmental disclosures resulted in findings similar to those of Cho et al. (2010). Specifically, our findings reiterated that low environmental performance results in the inclusion of ambiguous narrative content in environmental disclosures. However, whereas Cho et al. (2010) used American 10-K reports (which consist of financial reports for investors and shareholders in addition to environmental information) as data, we utilized sustainability reports issued by Japanese companies, which are intended for multiple types of stakeholders, such as customers, employees, and NGOs. In that respect, the behavior of manipulating narrative content to improve the company's image was observed. In contrast, the relationship between social performance and the narrative content of social disclosures could not be clarified. Hence, thorough examination of the factors that influence the narrative content of social disclosures remains a worthwhile endeavor for researchers in this area. In addition, it should be noted that our analyses utilize Japanese companies' sustainability reports in English, therefore proficiency in English needs not be taken into account.

Although the current study has left numerous avenues for future research, this paper represents the first instance in which the relationship between environmental and social performance and the sustainability reporting and the narrative content of those sustainability reports by Japanese companies has been empirically investigated.

Acknowledgements

Notes

[1] Reports related to environmental and social issues are commonly referred to as Environmental Reports, CSR Reports, or Sustainability Reports. Because they are voluntary forms of information disclosure, there is no standardized definition for them. For the purposes of the current research, annual reports related to environmental and social issues are called Sustainability Reports.

[2] Impression management theory relates to how to express oneself so as to improve others' perceptions. This theoretical perspective has been popular with researchers in social psychology (Felson, 1978).

[3] The professional services firm of KLD Research and Analytics, Inc. KLD was acquired by RiskMetrics in 2009, and RiskMetrics was acquired by MSCI in 2010. MSCI reports worldwide stock price indices.

[4] As a consequence of the popularization of the Environmental Report Guidelines promoted by the Ministry of the Environment, a number of Japanese companies publish an "Environmental Report" containing information related to their environmental performance. Although Environmental Reports will gradually transition to Sustainability Reports to comply with the Global Reporting Guideline, the present paper includes analyses of environmental reports, which are predecessors to sustainability reports.

[5] Corporate web sites also contain information related to the environment and society, but that some information includes topics that are not updated on an annual basis. For instance, environment management systems and CSR policies published by companies often include contents that do not need to be updated for a certain period.

[6] Social activities in Japanese companies chiefly consist of cleaning areas around factories, reforestation, and donations. Some large companies publish a leaflet that describes their social activities.

7 In GRI (2006), companies are required to disclose information related to community-directed investment which promotes the community's economic development and consequent independence. This deviates from the contents reported by many Japanese companies which principally consist of cleaning, reforestation, and donations.

[8] DICTION 6.0 allows four types of settings. These settings include: (1) "Abbreviated" (analysis of just the first 500 words), (2) "Averaged" (average analysis results of 500-word units), (3) "Unsegmented Average" (scores calculated for all sentences and normalized for each 500-word unit), and (4) "Raw Scores" (non-normalized scores).

[9] Other optional settings were the seven main categories: business, daily life, entertainment, journalism, culture, politics, and science. The "business" category includes seven types: "financial report", "public relations", "financial news", "legal documents", "magazine advertisement", and "TV commercials". For the purposes of current research, we selected the "financial report" item of the "business" category. The programming of "financial report" was based on financial reports from 48 Fortune 500 companies. The "business" category includes seven types "public relations", "financial news", "legal documents", "magazines", "magazine advertisement", and "TV commercials".

[10] STOXX is one of the world's leading independent index companies. STOXX's indices are provided to the world's largest financial and asset management companies.

[11] The ESG Leader Index includes the world's leading companies in terms of environmental, social, and governmental performance. ESG data of target companies are based on survey carried out by Sustainalytics. The index model was developed by STOXX.

[12] The ESG survey conducted by Sustainalytics is based on DVFA(Deutsche Vereinigung für Finanzanalyse und Asset Management) (2010) "KPIs for ESG – Key Performance Indicators for Environmental, Social and Governance Issues," a report that was jointly prepared with EFFAS. This report accounts for the viewpoints of investors, analysts, and rating agencies, and a KPI that can be used by those interested to perform evaluation is defined.

[13] Many companies in Japan do not disclose the social dimension in sustainability reports.

[14] Every company that consults the GRI has its database registered at the GRI web site. We considered those listed in that database as having followed the GRI guidelines.

References

Cho, H. G., Roberts, R. W. and Patten, D. M. (2010), "The language of US corporate environmental disclosure", *Accounting Organizations and Society*, Vol. 35, pp. 431–443.

Demers, E. A., and Vega, C. (2010), "Soft information in earnings announcement: News or Noise? ", *INSEAD Working Paper* No. 2010/33/AC, available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1153450, (Accessed 20 August 2012).

Deutsche Vereinigung für Finanzanalyse und Asset Management, (2010), "KPIs for ESG – Key Performance Indicators for environmental, social and governance Issues, " Available at<http://ec.europa.eu/enterprise/policies/sustainable-business/corporate-social-responsibility/reporting-disclosure/swedish-presidency/files/documents_left/kpis_for_esg_-_european_federation_of_financial_analysts_societies_en.pdf>, (Accessed 16 September 2011).

Felson, R. B. (1978), "Aggression as impression management", *Social Psychology*, Vol. 41, pp. 205-213, available at: http://www.jstor.org/stable/10.2307/3033557, (Accessed 22 August 2012).

Hart, R. P., and Carroll, E. C. (2010), "DICTION 6.0: The Text Analysis Program", Help Manual, Digitext, Inc.

Higashida, A., Kokubu, K. and Kawahara, C. (2005), "Information disclosure and determinants in environmental reports by Japanese companies: Analysis focused mostly on environmental reports issued in 2003", *Corporate Social Accounting and Reporting Research*, Vol. 17, pp. 39-38 (in Japanese).

Hooks, J. and Stadenb, C. J. (2010), "Evaluating environmental disclosures: The relationship between quality and extent measures", *The British Accounting Review*, Vol. 43, pp. 200-213.

Jeremy, C. S. and Timothy, B. P. (2007), "The application of DICTION to content analysis research in strategic management", *Organizational Research Methods*, Vol. 11, pp. 727-752.

Kida, M. (2006), "Cognitive study on Asahi Beer organizational reform: Text mining of negotiable securities", *Organizational Science*, Vol. 39, pp. 79-92 (in Japanese).

Kitora, Y. and Okuda, M. (2009), "The relationship between the basic corporate social responsibility (CSR) policy and corporate governance: A text mining approach", *Hitotsubashi Business Review*, Summer, pp. 152-163 (in Japanese).

Kokubu, K., Nishitani, K., Shinohara, A. and Kitada, H. (2012), "Environmental disclosure by Japanese companies: Influence of stakeholders and information needs", *Industrial Accounting*, Vol. 71, pp. 51-61 (in Japanese).

Kokubu, K., Noda, A., Onishi, Y., Shinabe, T. and Higashida, A. (2002), "Determinants of environmental information disclosure by Japanese companies: Issuance of environmental reports and qualitative analysis", *Accounting*, Vol. 54, pp. 74-80 (in Japanese).

KPMG International. (2011), "KPMG international survey of corporate responsibility reporting 2011", Available at: http://www.kpmg.com/Global/en/IssuesAndInsights/ArticlesPublications/corporate-responsibility/Documents/2011-survey.pdf, Accessed 28 October 2011.

Leszczynska, A. (2012), "Towards shareholders' value: an analysis of sustainability reports", *Industrial Management & Data Systems*, Vol. 11, pp. 6-12.

Mallin, C. and Michelon, G. (2012), "Monitoring intensity and stakeholders' orientation: How does governance affect social and environmental disclosure?", *Journal of Business Ethics*, Published online: 26 April 2012, DOI: 10.1007/s10551-012-1324-4.

Merkl-Davies, D. M. and Brennan, N. M. (2007), "Discretionary disclosure strategies in corporate narratives: Incremental information or impression management?", *Journal of Accounting Literature,* Vol. 26, pp. 116–194.

Neu, D., Warsame, H. and Pedwell, K. (1998), "Managing public impressions: Environmental disclosures in annual reports", *Accounting Organizations and Society,* Vol. 23, NO.3, pp.265–282.

Oda, E. and Mitsuhashi, H. (2010), "Mission statements and firm performance: An empirical analysis using text mining", *Japan Academy of Management Philosophy*, Vol. 1-2, pp. 22-37 (in Japanese).

Shino, S. (2010), "Content analysis of responsibilities to employees in Japanese companies' CSR reports: What for, For whom, What to do", *Aichi Gakuin University Discussions, Business Research*, Vol. 50, pp. 325-348 (in Japanese).

Shirata, Y., Takeuchi, H. and Ogino, S. (2008), "Corporate evaluation analysis using text mining technology: Experimental study on bankrupt companies", *Business Analysis Association*, Vol. 25, pp. 40-47 (in Japanese).

Short, J. C. and Palmer, T. B. (2008), "The application of DICTION to content analysis research in strategic management", *Organizational Research Methods*, Vol.11, NO.4, pp. 727–752.

Skouloudis, A., Evangelinos, A. and Moraitis, S. (2012), "Accountability and stakeholder engagement in the airport industry: An assessment of airports' CSR reports", *Journal of Air Transport Management*, Vol.18, pp. 16-20.

Sydserff, R., and Weetman, P. (2002), "Developments in content analysis: A transitivity index and DICTION scores", *Accounting, Auditing and Accountability Journal*, Vol. 4, pp. 523–545.

Table 1. Statistical quantities related to narrative content

	N	Mean	Std. Dev.	Minimum	Maximum
Environmental_disclosure_optimism	164	51.783	1.673	46.840	60.700
Environmental_disclosure_certainty	164	52.833	1.968	47.870	60.480
Social_disclosure_optimism	153	54.362	2.152	48.470	61.020
Social_disclosure_certainty	153	53.128	1.871	48.150	57.990
Environmental rating	260	58.869	27.737	2.500	99.400
Social_rating	260	41.489	26.019	0.100	97.400
Governance rating	260	44.764	24.787	0.300	94.800
Advertisement expenses to net sales	260	0.016	0.022	0.000	0.212
Free float weight	260	0.126	0.088	0.005	0.484
Financial institutions shareholding ratio	260	0.327	0.108	0.050	0.518
Foreign investor shareholding ratio	260	0.277	0.125	0.003	0.816
Size of the firm (log)	260	13.629	1.039	7.913	16.189
ROA	260	0.069	0.060	−0.047	0.490
Leverage	260	0.513	0.209	0.058	0.906
Number of pages on environment	164	13.232	13.217	1	130
Number of pages on society	153	10.444	9.179	1	70
GRI dummy	260	0.300	0.459	0	1
Manufacturing industry dummy	260	0.646	0.479	0	1
Sustainability report issuance dummy	260	0.631	0.484	0	1
UNGC Membership	260	0.200	0.401	0	1
Company age (log)	260	4.013	0.706	1.099	4.828

Table 2. Estimation results

	Model (1) Heckit		Model (2) Probit/OLS		Model (3) Probit/OLS		Model (4) Probit/OLS	
	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
Selection models								
ERAT	0.028	0.005 ***	0.028	0.005 ***	−	−	−	−
SRAT	−	−	−	−	0.013	0.004 ***	0.013	0.004 ***
GRAT	0.002	0.004	0.002	0.004	0.006	0.004	0.006	0.004
AER	1.540	4.681	1.578	4.715	1.496	4.386	1.496	4.386
FFW	−1.463	1.286	−1.385	1.274	−1.303	1.219	−1.303	1.219
FIR	−0.392	0.975	−0.311	0.980	0.309	0.927	0.309	0.927
FOR	−0.450	0.855	−0.449	0.861	−0.532	0.828	−0.532	0.828
SIZE	0.274	0.154 *	0.275	0.152 *	0.434	0.170 ***	0.434	0.170 ***
ROA	1.834	1.929	1.791	1.942	0.513	1.999	0.513	1.999
LEV	0.731	0.618	0.772	0.616	0.259	0.568	0.259	0.568
GC	1.221	0.350 ***	1.194	0.355 ***	1.009	0.309 ***	1.009	0.309 ***
AGE	0.150	0.130	0.135	0.132	0.098	0.146	0.098	0.146
MAN	0.304	0.247	0.299	0.248	0.974	0.226 ***	0.974	0.226 ***
Constant term	−6.072	2.260 ***	−6.072	2.249 ***	−7.459	2.708 ***	−7.459	2.708 ***
Performance models								
ERAT	−0.013	0.006 **	0.013	0.008 *	−	−	−	−
SRAT	−	−	−	−	0.007	0.007	−0.005	0.006
GRAT	−0.013	0.005 ***	−0.012	0.007 *	0.005	0.009	0.004	0.007
AER	−1.586	6.410	−5.901	6.434	−1.224	10.395	2.880	8.230
FFW	1.939	1.672	−2.639	2.552	−0.049	2.445	0.899	2.378
FIR	−0.634	1.217	1.461	1.590	2.841	1.694 *	1.257	1.620
FOR	1.191	1.461	−2.320	1.724	−1.589	1.379	1.910	1.667
SIZE	−0.027	0.110	−0.269	0.179	−0.055	0.193	0.071	0.149
ROA	2.631	3.430	5.436	4.570	−3.396	6.478	−2.450	4.689
LEV	−1.737	0.875 **	0.713	1.159	−0.200	1.170	0.431	0.954
EPAGE	−0.005	0.007	0.014	0.011	−	−	−	−
SPAGE	−	−	−	−	−0.008	0.017	−0.006	0.017
GRI	0.630	0.287 **	0.184	0.312	−0.522	0.377	−0.062	0.319
MAN	−0.039	0.268	−0.345	0.418	−0.541	0.429	−0.384	0.398
Constant term	54.066	1.596 ***	56.079	2.209 ***	55.090	2.816 ***	51.324	2.286 ***
N (selection models)	260		260		260		260	
N (performance models)	164		164		153		153	
Wald test (p−value)	0.087		0.332		0.111		0.529	
Log pseudolikelihood	−417.522		−		−		−	
Pseudo R^2 (selection models)	−		0.353		0.257		0.257	
R^2 (performance models)	−		0.087		0.062		0.049	

Notes: In Heckit, selection and performance models are estimated simultaneously; in OLS, they are estimated independently. Only highly reliable results according to the Wald test are shown. The White standard error was used. ***, **, and * respectively represent 1%, 5%, and 10% significance level (H_0: x = 0).