

(Measuring) Readability: a critical look at sustainability reports through standard formulae and NLP

1. Abstract

This study characterises and problematises the language of corporate sustainability reporting along region, industry, genre and content lines by applying readability formulae and Natural Language Processing (NLP) to a manually assembled 2.75-million word corpus. Readability formulae reveal that, despite its wider readership, sustainability reporting is still a very difficult to read genre, one sometimes more difficult than financial reporting. Although we find no industry impact on readability, region does prove an important variable, with Australian reports significantly more complex than others. These results not only highlight the impact of legislative contexts but also language variety itself as an underexplored variable. Finally, the study lays bare some of the weaknesses of default readability formulae, which are unable to detect American reports' more active language or lower lexical density in UK/EU reports, and demonstrates the merits of NLP in report readability analysis as well as the need for more accessible sustainability reporting.

2. Introduction

Sustainability reporting, under its various names and in its various forms, has increased considerably in adoption in the 21st century. The 2013 KPMG Survey of Corporate Responsibility Reporting indicated that 51 percent of companies issuing reports worldwide include Corporate Responsibility (CR) information in their annual financial reporting, up from 20 percent in 2011 and a mere 9 percent in 2008. KPMG sets a low barrier for entry in not specifying a minimum required amount of sustainability-related disclosureⁱ, but these numbers still reveal an undeniable trend. The survey reads that “the debate [whether or not to report] is over. [...] It is now about the quality of CR reporting and the best means to reach relevant audiences.”

Qualitative reporting delivers relevant information accessible to interested audiences, and one important facet of this process that scholars have devoted relatively little attention to is these reports’ readability. While sustainability reports have much in common with the notably difficult to read genre of corporate financial reporting (Courtis 1995, Courtis 1998, Stanton & Stanton 2002, Li 2008), the latter typically addresses a far more specialised readership of investors and analysts therefore better equipped to deal with textual complexity. The audiences that might benefit from corporate sustainability reporting, conversely, are remarkably diverse. Any stakeholder in the company’s operations, be they investor, employee or member of the community in which the company operates, might have an interest in its social or environmental performance. Not all members of this extended readership, however, will be able to decode a level of textual complexity similar to that of financial reporting. Even for the latter genre, Lehavy et al. (2011) show the importance of producing readable texts, demonstrating that investors will rely more heavily on expert analyses as a company’s reporting becomes less readable.

We are aware of only two previous studies into the readability of corporate sustainability reporting, both echoing often-conducted studies into the (poor) readability of corporate financial reporting. Farewell et al. (2014) voice concerns that “the average customer” will struggle to decode these reports, and conclude with the plea that “companies should work harder to choose simple language,” which reinforces KPMG’s demand for higher-quality reporting. Sheikh Abu Bakar and Ameer (2011) find consistently high reading difficulty across a sample of Malaysian CSR communications, noting that that their communications’ readability deteriorates as company performance does. These findings support the ‘obfuscation hypothesis’ (Courtis 1998, Rutherford 2003), which posits that companies will make unfavourable news more difficult to decode. Research on the use of visuals in CSR offers further evidence for obfuscation and impression management. Cho et al. (2012a, b) found that sustainability reports, just like financial reports, show a preference for graphs that display positive trends while additional graph distortion is used to embellish results. Studies like Hrasky (2012) and Boiral (2013) in their turn illustrate how attractive imagery is used for window-dressing and green-washing in those cases where the impact of sustainability measures is unclear. These reports’ susceptibility to manipulation of their presentation also further justifies examining their textual content.

We wish expand research into corporate sustainability reporting readability through expansion in scope, genre and variety, and means of analysis. Our approach yields three key differences with previous studies. First, we measure the impact of language variety and industry by examining readability across a larger corpus than previous studies which consists of approximately 2.75 million words, representing five language varieties and four industries, totalling 470 texts. Second, by using a genre-diversified corpus we aim to compare sustainability-related content’s readability compared

to financial content and verify whether the former takes its extended audience's accessibility requirements into account compared to financial reporting. For those purposes, we included not only sustainability-related disclosures in the corpus, but also the same companies' chairman's letters (the most often-consulted sections of corporate reports according to Curtis 1998, Clatworthy & Jones 2003 and others) for both the company's annual financial and sustainability reports. Third, we expand on previous research by using Natural Language Processing tools for a finer-grained level of analysis in addition to the often-employed 'shallow' readability formulae. In addition to measuring readability in terms of word and sentence length, NLP tools can quantify e.g. the use of passive structures, the syntactic (and thus, potentially, relational) depth of a given sentence, or lexical density.

This expansion of scope and methodology compared to previous research not only allows us to describe sustainability reporting's language in greater detail, but examine whether purely linguistic features of prescriptive regulations or the language variety itself (see Precht 2003a, for instance, on American directnessⁱⁱ) aid in characterising the genre's complexity, similar to how extent of legal enforcement can aid in predicting extent of earnings management (see Leuz et al. 2003). In addition, it induces a more critical stance towards the entrenched use of readability formulae and opens up avenues for more in-depth analysis with more nuanced perspectives on reports' use of language and its impact across borders and language barriers in an increasingly internationally prominent field.

The rest of this article is structured as follows. First, we present a literature review on readability, language variety and linguistic complexity in English, and sustainability (reporting) and its altered readability requirements. Second, we will discuss the corpus and how we compiled it. Third, we will examine the readability of the corpus, formulate a number of hypotheses and test the impact of a number of variables such as industry and language variety, and expand on traditional readability analyses through Natural Language Processing techniques. Finally, we present our conclusions along with avenues for future research.

3. Literature Review

3.1. Readability

Scholars have fiercely contested the merits and methods of expressing a document's readability numerically ever since "A new readability yardstick", Rudolf Flesch's seminal (1948) foray into the field. Edgar Dale and Jeanne Chall (1948), Robert Gunning (1952) and J. Peter Kincaid (1975) devised a few of the better-known formulae for readability calculation still employed today, but the formulae proposed for readability calculation numbered in the hundreds by 1980, and that of its advocates and critics in the thousands (DuBay 2004). However, consensus on how researchers should define readability remains elusive. DuBay (*ibid.*), for example, defines it as the text-internal characteristic of "what makes some texts easier to read than others", opposed to the formal aspect of legibility, "which concerns typeface and layout," but emphasises 'understandability' and 'comprehensibility' as concepts crucial to readability in most definitions he cites (e.g. Klare 1963, McLaughlin 1969). Conversely, Smith and Taffler (1992) oppose readability with understandability. They describe the former as a set of purely text-internal characteristics that determine difficulty and the latter as the interaction between the text and its reader, with prior knowledge also affecting comprehension.

Previous studies into the readability of corporate reporting, typically the annual financial report (e.g. Courtis 1995, 1998, Li 2008, Lehavy et al. 2010 and Kumar 2014), consistently portray it as a far more difficult genre to read than the average text. Jones and Shoemaker (1994) and Courtis (1995) also note the potential impact of the intended audience's prior knowledge. The latter describes the prose in annual reports as "beyond the fluent comprehension skills of about 90 per cent of the adult population and 40 per cent of the investor population" based on findings derived from readability formulae. Courtis suggests a variance in sophistication between different investors, but also implies that they are the primary intended audience of financial reporting. Financial motives will similarly drive most other groups interested in these financial reports, e.g. analysts. Although many readers of sustainability reporting will similarly consult these reports to make informed investment decisions (see e.g. Carnevale & Mazzuca 2014), section 2.3 expands on how these report also appeal to an extended stakeholder audience with different readability requirements.

As the concept of readability remains contested, this study will only assume that when a text's features make it easier for the reader to extract the information they want, it is more readable. Most changes in the text that make it easier for a reader unfamiliar with the genre to gain such information will make that same process easier for the sophisticated reader, even though the ultimate extent of their understanding may differ acutely. For example, the Securities and Exchange Commission (1998) promotes active over passive voice in order to facilitate reading. Such a change might make a fairly accessible text straightforwardly accessible to an expert, and simplify a non-expert's experience from very difficult to merely difficult.

Our definition of readability is also purely text-internal. Paratextual features such as font, layout, pictures and graphs almost certainly impact most readers' interpretation of these reports, and are also susceptible to obfuscating manipulation (see Cho et al. 2012a, b). However, as the technical limitations of our means of analysis demands that we discard this paratext, we do not integrate it into our definition.

Finally, while we do rely on the familiar readability formulae, we will also echo the familiar caveat that they are a cost-effective means of estimating readability but do not necessarily allow for a fine-

grained or universally applicable analysis of how easy a text is to decode. We will explore how NLP might assist in fine-grained text analysis.

Our analyses use three common readability indices: the Flesch Reading Ease Score, the Flesch-Kincaid Grade Level score, and the Gunning Fog Index. The Flesch Reading Ease Score (Flesch 1948) is one of the oldest and still most widely-used formulae for computing readability and is thus most suited for the first step of our inquiry, i.e. comparing our corpus' readability with that of other genres. Many scholars, such as Curtis (1995) adopt Flesch's 1949 expansion on his original work of defined 'degrees' of readability from 0-100, with a range of 0-30 signifying the lowest reading ease and incrementing in steps of ten from there on. Drawing on the same textual variables as the Flesch Reading Ease Score (Kincaid 1975), the Flesch-Kincaid Grade Level attempts to quantify the years of education that the text requires of the reader. While the above caution that these automatic formula are approximations at best certainly applies here, grade levels allow for more intuitive results than the Flesch Reading Ease Score does. The Gunning Fog Index (Gunning 1952, revised in Bogert 1985), finally, attempts to distil a grade-level measure of readability like the Flesch-Kincaid formula does, but places a stronger emphasis on the ratio of polysyllabic ('complex') to mono- or disyllabic words present in the text. Studies into annual report readability often measure it through the Fog Index (e.g. Li 2008, Leheavy 2011), instead of or in addition to Flesch-based readability. By incorporating the Fog Index into our metrics, we can straightforwardly compare our results with those of previous studies into corporate report readability. Table 1 contains the exact formulae we used to calculate these readability measures.

TABLE 1

We also expand on these formulae by to quantifying the deeper-level linguistic features of lexical density, subordination, parse tree depth and passivisation (section 4.5). Lexical density quantifies the number of content words (e.g. 'sustainability' or 'company') relative to the number of grammatical words (e.g. 'if', 'but', 'will'). Higher lexical density can lead to higher textual complexity (Halliday 1989, Harrison & Bakker 1998) due to a higher conceptual load. We quantify subordination as the average number of subclause-introducing elements per sentence, which serves as a syntactic complexity measure (Beaman 1984, Dell'Orletta et al. 2014), and parse tree depth as the average number of levels in a sentence parse tree, which can indicate complexity and cognitive load (1964, Dell'Orletta et al. 2014) Finally, we measure the average number of passive structures per sentence, which the SEC advises against in its Plain English guidelines (Securities and Exchange Commission 1998)

Figure 1. Sample CoreNLP parse (Stanford NLP Group 2015) for the sentence 'In fact, much of the water used in production is of such good quality that we have official approval from the relevant authorities to discharge it directly into rivers.' (Infineon 2013)

FIGURE 1

A deep-level syntactic analysis is impossible without a parsed corpus and manually parsing a multi-million word corpus is not viable, and infeasible without Natural Language Processing technology capable of automatically analysing text. NLP tools are significantly more technically demanding to implement than readability formulae, but do allow for finer-grained analysis. The parse tree above, for instance, visualises the levels of syntactic depth, with elements to the right deeper in the tree, and signals an instance of subordination as 'SBAR', whereas traditional formulae would only measure sentence and word length.

3.2. Language Variety

Scholars have hitherto largely neglected the impact of language variety on corporate reporting, or, in the best case, seldom explicitly isolated it as a variable. For instance, Leuz et al. (2003) distinguish between three clusters of declining legal enforcement: the US, the UK and Australia belong to the cluster with the highest enforcement, most European countries (save Greece, Portugal, Italy and Spain) to the second, and the remaining ones, along with India, to the last, which faces the least legal enforcement. As the study finds that clusters with greater legal enforcement exhibit less earnings management, we might similarly expect that the countries in the first cluster will exhibit less textual manipulation, and thus better readability, than those in the clusters with a lesser extent of enforcement. Cho et al. (2012b), also drawing on Leuz et al.'s framework, find a greater skew towards positive graphs in countries from less-regulated clusters, similarly suggesting manipulation. Language variety is present by proxy in this analysis as we mainly find those countries with English as a sole official language in the first cluster and countries that employ Business English as a Lingua Franca (BELF) in the second and third, which more linguistically diverse India also occupies.

Scholars such as Precht (2003b) and Creese (1991) do suggest potentially relevant variation between varieties in the same cluster in their application of such syntactic and semantic elements as passivisation, impersonalisation and directness. As corporate reports reach ever-increasingly international audiences (Townsend et al. 2012), we wish to examine how textual complexity, expressed both as a 'shallow' formula or a set of linguistic features, differs across the five varieties present in our corpus. For instance, a British report might contain more passive structures in order to express itself less directly and maintain an (expected) discursive distance from the British reader, but might, in doing so, strike an American reader as evasive. This analysis may demonstrate that, beyond the different clusters of institutional climates that might impact reporting, the different varieties of English also represent different linguistic attitudes that may influence report readability. Results that indicate such differences may spur future research to consider a mainly linguistic approach to language variety as an important variable in its own right.

3.3. Sustainability

Corporate sustainability is as contested a concept as readability is, with a plethora of terms, prime among which 'sustainability' or 'corporate (social) responsibility', that both scholars and producers as well as users of these reports use to describe various, often overlapping concepts. Dahlsrud (2008) counts a non-exhaustive 37 definitions, cautiously reiterating Van Marrewijk's (2003) warning that bias towards specific definitions stagnates the concept of CSR and risks opportunistic selection. While companies' definitions of corporate sustainability can vary greatly, its implementation sees more uniformity due to the trans-nationally employed Global Reporting Initiative guidelines' (2013) position as the de facto standard for sustainability reports' form and content (Temouri & Jones 2014).

According to Dahlsrud, the Commission of the European Communities' (2001) definition of Corporate Social Responsibility sees most use. It presents CSR as "[a] concept whereby companies integrate social and environmental concerns in their business operations and in their interaction with stakeholders on a voluntary basis." Dahlsrud notes the completeness of this definition, which encompasses five terms key to CSR: 'economic', 'voluntariness', 'stakeholder', 'social' and 'environmental'. If we widen our scope beyond the corporate, the World Commission on Environment and Development ('Brundtland Commission')'s seminal (Hahn & Kühnen 2013) 1987 report *Our Common Future* offers the rhetorically salient definition of sustainable development as "meet[ing] the needs of the present without compromising the ability of future generations to meet

their own needs,” a phrasing that still echoes throughout corporate sustainability reports (e.g. Infineon 2013).

As previously mentioned, we expect that sustainability- or CSR-related reporting has the potential to appeal to a wider audience than strictly financial reporting. Typically, the latter will primarily be relevant for shareholders. The former can be relevant to a company’s stakeholders in the widest sense. This distinction draws on Sacconi (2004), who presents stakeholders in the strict sense as “those who have an interest at stake because they have made specific investments in the first (in the form of [human, financial, social, physical or environmental capital, or trust])”, opposing them to the wider-sense stakeholders, whom corporate operations impact positively or negatively even without their direct participation. Sacconi consequently models Corporate Social Responsibility as a model of corporate governance in which the company should not just represent the interests of its owner(s), but those of “*all* its stakeholders (the owners included).” Companies adhering to this model of CSR may very well wish to address a maximal proportion of this wider sphere of stakeholders in their reporting, be it out of reputation management concerns (ibid., Bebbington 2008), moral obligation, or a desire to legitimate corporate operations (Tilling 2004).

This sentiment echoes throughout many sustainability reports. Virtually all of them acknowledge these ‘wider-sense’ stakeholders, such as the communities in which companies operate, as integral to CSR, and commonly acknowledge their position in these reports’ readership. For instance, mining company Lonmin states in its first sustainability report (2013) after what it refers to as ‘The Marikana Incident,’ⁱⁱⁱ which led to 46 deaths, that “[t]he decision to produce a printed report was made to ensure accessibility by more stakeholders who may not have access to the report on an online platform.” Total (2013) points out its efforts to submit its CSR report to wider-circle stakeholders such as NGOs or governments local to operations and gather feedback. The Adidas Group’s (2013) sustainability reports welcomes their reader to the report with the acknowledgement that “you and many of our consumers and stakeholders have high expectations [...] when it comes to [our] sustainability efforts”, suggesting a wide intended audience for the report. Not all reports mention their intended readership explicitly, but many integrate the breadth of their CSR efforts into their corporate sustainability narrative, acknowledging the widened circle of stakeholders beyond shareholders and partners in the process.

These reports’ readers also appear to profile themselves as consumers more than investors in their use of sustainability reporting. Townsend et al.’s 2010 Readers & Reporters [of Sustainability Reporting] Survey indicates that the primary motivation behind readers’ use of sustainability reports is ‘inform[ing] decisions on use of the organisation’s products/services’, closely followed by ‘inform[ing] investment/divestment decisions’. While the investment motive remains prominent, sustainability and financial reporting audiences’ motives do appear to diverge. The authors define ‘readers’ rather broadly, as ‘any stakeholders that have been engaged with an organisation’s reporting output’. The survey indicates the average reader read three reports, but the top five percent of readers read between ten and twenty reports per year.

While the authors’ sampling mechanism is somewhat unclear, we observe that, at least for this sample, most readers are non-expert, possibly casual users of these reports likely to lack the sophistication of veteran analysts or investors that regularly read financial reports. The composition of the sample also suggests a notable expansion beyond the likely audience for financial reporting. While 48% of the sampled readers were company-internal and 16% were investors, 14% consisted of the company-external value chain and, finally, 22% originated from ‘Civil Society’, entailing media, labour unions, public institutions, academics and other experts, and concerned citizens and consumers. Given the diversity of the audience, we might expect companies to adjust the

accessibility of their reports accordingly by reducing informational complexity, e.g. by ensuring high readability compared to financial reporting, as section 4.3 will examine in greater detail.

Neither scholars nor the companies themselves portray corporate social responsibility and corporate sustainability as purely altruistic concepts, despite the lofty foundations set out above. Scholars such as Parsons & McKenna (2005) and Boiral (2013) fiercely criticise the extent to which the language used in sustainability reporting twists its narrative frames to the company's advantage^{iv}. Story & Neves (2015) point out the risks of alienating readers when they perceive corporate social responsibility initiatives as purely strategic. The reader's perspective appears less critical, however^v: Townsend et al.'s (2010) survey indicates that a minority of readers still sees sustainability reporting as 'greenwashing' (the 2013 KPMG report points out this same shift in attitude) and they consider corporate accountability the prime motivation behind reporting.

Nevertheless, researchers do acknowledge CSR is a prime means of corporate reputation and impression management (e.g. Sacconi 2004 and Cho & Patten 2012, resp.) and companies often note the financial bottom line (i.e. profit) as a crucial element of their 'triple bottom line' (Elkington 1997) or variants upon that term. The first notion is linked to the more favourable reputation companies stand to gain from both their social and environmental initiatives and general conduct. Companies' own narratives often posit that sustainable operation, and thus sustainable profit, entails operating in such a way that the company maintains its social and environmental licence to operate (Deegan et al. 2002). They can also benefit from the added (perception of) transparency stemming from the sheer act of disclosing non-financial information. Ideally (but not always), this is done with a balance between positive and negative news as proposed in, for example, the Global Reporting Initiative guidelines' (2013) tenet of 'balance', which Section 1 already anticipated as often problematic in both textual and non-textual content.

In summary, CSR is a mosaic of concepts, and any study will struggle to distil them into a single definition. Fortunately, as was the case for readability, this study can cast a wide net: for most purposes, we do not need to define sustainability, CSR or its related terms beyond how the authors of the reports in the corpus define them, as the company's individual definition ultimately determines which content makes it into the report. Regardless of reporters' specific interpretations, this enables us to analyse a genre which, like the financial report before it, demands inquiries into its readability because of its meteoric rise in prominence, but has received little such attention. As is still the case for the financial report, both producers and readers of sustainability reports stand to benefit from having disclosures readable and accessible to their intended audience, because to do otherwise is typically a waste of resources, an attempt at obfuscation, or both. That makes the sustainability report's wider audience of stakeholders a crucial determinant in how such reporting should be conducted, albeit one that Section 4's results suggest may often be ignored.

4. Methodology

4.1. Corpus

Our corpus consists of 470 texts, split along three primary axes: region, industry, and (sub-)genre. These texts represent approximately 3.95 million tokens, 2.75 million of which running plain text usable for NLP purposes^{vi}. The five regions present in the corpus are the United States, the United Kingdom, (non-UK) Europe, Australia, and India. The four industries are mining and metals, oil, semiconductors and apparel. The three subgenres present are the sustainability report and two types of CEO or chairman's letter or address, one type oriented towards financial performance and the other concerning sustainability. Table 2 shows the number of texts per region and per industry for each of the genres.

TABLE 2

We selected these four industries in order to achieve a diversified selection of potentially environmentally sensitive (cfr. Cho and Pattern 2007) companies in the cases of Mining/Metals and Oil, and potentially socially sensitive companies in the cases of Semiconductors and Apparel. We assert the latter two's sensitivity to social issues on the basis of the use of conflict materials (Bafilemba et al. 2014) and recent controversies around worker rights (e.g. the Savar building collapse, BBC 2013), respectively. We selected five varieties of English to achieve a balance between some of the most prominent ones used in business today. The three (sub-)genres, finally, isolate the most often-read component of corporate reporting (Courtis 1997), the CEO letter^{vii}, for both types of content. The separation from the rest of the sustainability report proper serves to reveal differences between this most-consulted section and the more in-depth reporting narrative.

However, several of these cross-sections lack entries altogether, and the totals indicate a skew towards the extractive industries (mining and oil). While corpus balance and representativeness may seem suboptimal based on these numbers, the data at least partially satisfy the latter criterion as this is the complete set of texts available based on the data collection process set out in the next section.

4.2. Data Collection and Processing

This study is a first step in a larger project that aims to discover how linguistic and both financial and non-financial performance information interact in corporate reporting. No corpus was readily available, and our aims placed a significant restriction on the creation of one: for every text collected, we needed both financial and non-financial performance data. It proved more efficient to start from a financial database that also contains environmental, social and governance (ESG) information and collect texts for those companies for which both types of information were available, in order to enable future analyses of the relationship between performance and language. Thomson Reuters Datastream offered both a summary performance score and many discrete data points such as 'Sales Revenue' or 'Greenhouse Gas Emissions Volume' through its ASSET4 database^{viii} (Thomson Reuters 2013). We selected companies for the corpus based on Thomson Reuters' 'ASSET4 Template', which contains a list of all companies present in the 'ASSET4 Universe,' i.e. companies with information available, divided by industry and country. We adopted ASSET4's industry and country divisions and collected available texts for all listed companies within 'Mining/Metals', 'Oil', 'Textiles/Apparel' and 'Semiconductors' for the United States, United Kingdom, Australia, India and all non-UK European countries.

We downloaded, where available, the annual financial and sustainability reports for fiscal 2012^{ix} from the corporate website of all companies included in ASSET4 that met our criteria. When

available, we integrated the entirety of the company's separately published sustainability report into the corpus, only separating CEO letters into a separate document. We consider CEO letters a separate sub-genre as they are the most-read report section, addressing the reader directly through its own rhetoric.

When the company did not issue a separate sustainability report, we extracted any chapters with sustainability, CSR or ESG-related keywords such as 'health and safety' or 'community engagement' present in the heading from their financial annual report and classified them as the company's (de facto) sustainability report. Finally, for the vast majority of companies this annual financial report also contained a CEO letter or chairman's address (typically indicated as such in the table of contents) focusing on the company's financial performance. For a fair number of companies in our corpus, this is the only text type available out of the three.

We used ABBYY FineReader for conversion from the report PDFs to plaintext usable for our analysis. During this conversion, we manually tagged any numerical or mixed-content tables encountered in order to be able to remove them and pass only running text along to the NLP toolset. A regular expression^x stripped away all paragraphs that did not adhere to sentence case and end in a period, colon, semicolon, question mark or exclamation mark. This extracted running text and purely textual enumerations, removing, for example, headings, subheadings, and tables. Finally, we normalised case so that any word starting with a lowercase character was changed entirely to lowercase in order to reduce OCR casing errors. This process discarded about 1.2 million tokens, conserving 2.75 million out of an initial 3.95 million. To ensure sufficient remaining text length, we eliminated those texts less than 200 words long after cleaning from our analysis.

After extracting the running text, we analysed each of the trimmed files using the Stanford CoreNLP suite (Manning et al. 2014), which automatically annotates its input for part-of-speech, presence of named entities, syntactic structure, coreference, and other linguistic features. This allows us to quantify key linguistic aspects of our corpus not just in terms of readability formulae, but also deeper-level syntactic parameters such as the average depth (in levels) of the parse tree, the amount of passive structures per sentence, or use of various types of connectives.

5. Analysis and Discussion

Scholars have been investigating corporate reporting through readability measures for decades, but have conducted few such studies on the corporate sustainability report, likely due to the relative novelty of the (sub-)genre. Readability analysis of various genres through abstract formulae is a sufficiently long-standing practice to be a tradition, but continuous refinement of Natural Language Processing techniques now also facilitates the analysis of more complex linguistic phenomena that might help predict readability.

Companies' own narratives and previous studies suggest that corporate sustainability reporting has an extended stakeholder audience compared to financial reporting's primarily shareholder audience. Based on that discrepancy, we formulate the following hypotheses:

H1: sustainability information in standalone sustainability reports will be more readable than that in financial reports.

H2: corporate sustainability reports will be more readable than financial reports.

Based on previous studies into the effect of language variety on syntactic structures and different regions also representing different clusters of legal enforcement, which influences earnings management, we formulate the following:

H3: report readability will differ by region.

H4: reports from clusters with higher legal enforcement will be more readable.

Finally, based on the assumption that environmental and socially sensitivity may alter reporting language, we formulate the following:

H5: report readability will differ by industry.

Before testing these hypotheses, we will characterise our corpus in terms of formula-based ‘shallow’ readability. In the process of testing them, we will explore the predictive power and added value of linguistic features over readability formulae.

We used SPSS version 22 in order to conduct analyses. These consist, except where noted otherwise, of univariate general linear models investigating the impact of one or several predictor variables on a single measure of or proxy for readability. A single data point represents a single text from the corpus. We conducted most analyses only on the sustainability reports proper, omitting both types of CEO letters except where noted otherwise. We set the α level at .05 and apply Bonferroni correction for post-hoc analysis.

TABLE 3

TABLE 4

5.1. General readability statistics

Virtually all scholars (see sections 3.1 and 5.3) who have conducted this type of analysis agree in their findings that the corporate annual report is a difficult – often very difficult – genre to read. The few studies conducted into sustainability report readability find the same, observing little to no positive impact from the change in subject matter or potential widening in audience. This corpus proves no exception, with the sustainability reports having a mean Flesch score of 16.05 and a standard deviation of 7.78. This positions the sustainability report positions within the most difficult spectrum (0-30) that the Flesch Reading Ease Score distinguishes. The Flesch-Kincaid Grade Level mean stands at 16.85 with a standard deviation of 1.87, suggesting that these reports’ textual complexity is well beyond the comprehension of a reader with secondary school education, and likely most accessible for holders of a graduate degree.

The Fog Index, similarly intended to quantify the years of formal education necessary to comprehend a text, has a mean of 21.32 with a standard deviation of 2.12. If we expand this initial analysis to include the full corpus, i.e. both types of CEO letters as well as sustainability reports, we find a slight reduction in overall textual complexity, with a mean Flesch score of 21.44 (SD 9.40), Flesch-Kincaid Grade Level mean of 16.25 (SD 2.02) and Fog Index mean of 20.62 (SD 2.26), which is extremely, perhaps disproportionately large in suggesting readers might need eight or more years of higher education in order to fully decode these reports^{xi}. In summary, all three formulae point towards a remarkably high level of reading difficulty for sustainability reports and corporate reporting in general consistent with previous studies.

5.2. Standalone vs. financial report-internal sustainability disclosures

As we extracted sustainability disclosures from both annual reports and standalone sustainability reports, we wish to analyse whether there is any difference between the readability of sustainability information in financial reports and that in standalone sustainability reports. We compare the means and conduct Analysis of Variance (ANOVA) for both types of report for the three readability formulae.

We find some evidence for H1, observing a generally significant (Flesch score $p = 0.53$, i.e. narrowly insignificant; Kincaid score $p = 0.014$; Fog Index $p = 0.006$) difference between sustainability content present in financial reports and standalone sustainability reports, albeit with small effect sizes for each of the formulae. These results suggest more accessible writing for content composed specifically for the broader audience that sustainability reporting may want to address, compared to sustainability content they intend specifically for shareholders. Nevertheless, the small effect sizes render it doubtful whether readers will notice any difference. Regardless of the origins of the content, its mean readability scores still mark it as very difficult to read.

Given the small effect sizes, this difference between the two origins for sustainability content should not noticeably affect the following analyses, especially given the relative balance of both types of sustainability content within the corpus, although we do note that if we only drew on standalone sustainability reports, the corpus would be slightly more readable overall.

5.3. Financial vs. Sustainability Content

Courtis' (1995) overview of Flesch Reading Index score means for various US, Canadian, New Zealand and Hong Kong studies collects a set of Flesch scores from studies on Chairmen's addresses, with a lower bound of 28.96 and upper bound of 47.2. Flesch scores from financially oriented CEO letters from our corpus still score significantly ($p < 0.001$) below Courtis' lower bound with a mean difference of 2.90, which aligns with Jones and Shoemaker's (1994) findings that the mean textual complexity of corporate reporting may increase outright over time. Nevertheless, we note the small effect (Cohen's d of 0.36) for the difference in time compared to the difference in theme (large d of 0.95). Part of the increase in complexity for sustainability reports in our corpus is likely due to the hypothesised increase in complexity over time across all types of reporting, but most of it is likely due to report writers using more complex language for sustainability content than they do for financial content. We find additional evidence for this in our set of sustainability-related CEO letters having a significantly lower ($p < 0.001$) Flesch score than Courtis' lower bound for CEO letters with a mean difference of 9.16 (large d of 0.97), similar to previous comparisons. In summary, we find considerable evidence for corporate reporting as a genre becoming less readable when dealing with sustainability topics compared to financial ones. This is especially striking given that sustainability content in standalone reports is slightly more readable than that in financial reports, which potentially marks the sustainability sections in financial reporting as particularly unreadable passages.

Li's 2008 study into the readability of annual reports offers a sample Fog Index score mean for textual passages in (financial) annual reports of 19.4. A two-tailed t-test reveals that the sustainability reports from our own sample are significantly ($p < 0.001$) less readable, with a mean difference of 1.92 (large d of 0.90). Leavy et al. (2011) report a near-identical Fog mean of 19.53, again marking the sustainability reports as significantly ($p < 0.001$) less readable with a mean difference of 1.79 (large d of 0.84). Courtis' aforementioned 1995 study also compiles the results of studies into the readability of (financial) annual report footnotes, attesting means between 23.49 and 34.29. A two-tailed t-test again indicates significantly ($p < 0.001$) greater complexity for the sustainability reports, with a mean difference of 7.43 (large d of 0.95) with even the least readable of these annual report readability means. Contrasting the two types of CEO letter within our corpus yields a very similar result. In summary, the difficulty of text used in sustainability reporting appears to run contrary to what its potentially wider audience would suggest, thereby contradicting H2. According to our findings, the language used in sustainability reporting is structurally more complex than even annual financial reporting.

These means indicate the same trend towards sustainability-related content being less readable, with each genre differing significantly from the others in reading difficulty across the three measures. The Flesch score sees financially oriented CEO letters attain a significantly ($p < 0.001$) higher level of readability with a mean difference of 6.26. There appears to be a noticeable effect, with a medium, close to strong, effect size. The same is true for the Flesch-Kincaid score, albeit with a less pronounced significance ($p = 0.04$), effect size (small), and mean difference of 0.62. The same effect is present but no longer significant ($p = 0.13$) for the Fog score with a mean difference of 0.56. Jones and Shoemaker (1994) note that chairman's addresses are generally more readable than other sections of corporate reporting, suggesting why the order of most to least readable is consistent across the genres, if not always significantly or strongly different. Financially oriented CEO letters register as more accessible than sustainability-related CEO letters, which are more accessible than sustainability reports proper.

Both the sustainability report's poor general readability and its increased complexity compared to financial reporting reflect unfavourably on many companies' implicit or explicit desire to involve a greater number of stakeholders into the 'triple bottom line' reporting process. In spite of efforts to make these reports *available* (i.e. physically or digitally obtainable) for all stakeholders, sustainability-oriented disclosures are unlikely to be *accessible* (i.e. feasible to extract desired information from) to a potentially much wider sphere of stakeholders.

5.4. Language variety

We compute readability variability between varieties for the sustainability reports across the three primary formulae.

We find considerable evidence for H3 as across the three different measures, Australian English is consistently and significantly ($p \leq 0.003$) more complex than British and American English and European BELF, with medium effect sizes overall. The difference with Indian English is never significant, likely due to the relatively small sample size for the latter. This effect might be partially due to Australian Stock Exchange reporting requirements facing some flux between less stringency than US or UK requirements (CPA Australia 2013) and rising reporting standards (Nearmy 2014) at the time of data collection.

Nevertheless, we do indeed find differences within regions belonging to the same cluster within Leuz et al.'s framework but do not see salient deviations in readability between those different clusters. These results contradict H4, and in tandem with H3 evidence an often ignored language variety-internal influence on these reports' readability that goes beyond reporting climate, and in the subsequent section we will attempt to expand on existing research by exploring this discrepancy between varieties through an expanded, finer-grained linguistic framework.

5.5. Deeper-level linguistic features

To expand on the relatively shallow picture that these automated readability formulae paint, we examine the average amount of subordination used, the reports' average syntactic tree depth, the average number of passives per sentence, and the reports' lexical density. Subordination indicates how many linguistic elements, on average, embed sentences into other sentences, and syntactic tree depth expresses the average number of such levels per sentence. Both can thus measure relational complexity. Passive structures can complicate text by obscuring agency, and lexical density indicates how many content words a text contains compared to function words. While all of these 'deeper' features might potentially correlate with the various readability formulae by influencing sentence length, the readability formulae are, by design, unable to measure them directly.

Australian sustainability reports show a consistently higher mean amount of subordination than other varieties, but the difference is only significant ($p = 0.017$) for the Indian reports, despite the low sample size for Indian reports. Similarly, Indian reports appear to employ less subordination than other varieties, but the difference is not significant. The small effect size marks these effects as potentially evading notice. Australian reports also have consistently deeper parse trees than the other varieties, significantly ($p = 0.006$) so for the US, with the difference with Europe and India only approaching the threshold of significance ($p = 0.110$ and 0.062 , respectively). As before, effect size is small.

Other linguistic variables reveal more than the 'shallow' formulae could. Australian reports employ significantly ($p \leq 0.031$; strong effect) more passivisation than all varieties but Indian English, again likely due to its small presence in the corpus. Additionally, we find remarkably strong evidence for Precht's (2003a; 2003b) assessment of American English as more active and direct: the American reports use significantly ($p \leq 0.048$) fewer passive structures than the other language varieties. This variation sees a strong effect, indicating that these reports' readers may well notice the more active language. While the latter is less complex, we find no formula-based evidence of significantly lower complexity for the American reports in spite of their more active language. We note that increased passivisation can serve as an obfuscation strategy by concealing agency and, thus, responsibility, and that the influential (Colesanti 2012) US Securities and Exchange Commission Plain English Reporting Guidelines (1998) recommended against passive structures, which might encourage active voice in US reports regardless of linguistic attitudes.

Finally, British reports display a significantly ($p = .022$) lower lexical density (i.e. relative number of content words) than the other regions, Europe excepted, which is also significantly less lexically dense ($p < .001$) than India and the US, with a medium effect size on the threshold of a strong one. As was the case for American directness, this effect might tie into one of the variation stereotypes, i.e. the previously asserted (Precht 2003a, 2003b) British tendency towards modality and hedging, which might vex an American reader. While this causation would be very difficult to ascertain without manual filtering or counting, we formulate this conjecture based on the notion that modal verb structures can contain additional function words expressing modality, such as 'might yield results' versus 'yields results'. Precht (2003a) attests an increase in modal verb use relative to lexical verb use in spoken British compared to spoken American, which may lower lexical density. While a higher lexical density might indicate a higher conceptual load and thus lower reading ease, the readability formulae are unable to reflect that. Conversely, readability formulae might, contrary to intuition, register the increase in modal structures as a readability-lowering factor as it increases sentence length. Quantifying lexical density, like passivisation, paints a far more nuanced picture of these reports' language than the formulae might.

In summary, results for subordination and syntactic tree depth, two closely interlinked concepts, are similar to those for the readability scores, albeit finer-grained, which suggests effects for Australian and Indian English that were not visible to the formulae, but may be relevant to obfuscation analysis. The case of lexical density reveals additional effects pertinent to British English and European BELF, while passivisation both reinforces the formulae's findings Australian English and reveals additional effects for American English. In spite of these deeper syntactic variables not being directly visible to the automatic formulae, these results do demonstrate their relevance as they paint a similar picture of textual complexity between the different varieties – with Australian reports proving notably more complex across a set of different features – but nuance it with additional effects invisible to the formulae, also partially mollifying the extreme results the formulae sometimes indicate.

5.6. Industry effect

Apart from (sub-)genre and language variety, the corpus is also divisible along industry lines. Farewell et al.'s (2014) study, for instance, calls for the investigation of industry as a determinant of sustainability report readability, and one potential division that shaped the current corpus is between the industries that we might assess as more environmentally sensitive (mining and oil) and those we might under the same terms assess as more socially sensitive (semiconductors and apparel).

Overall, we find little support for H5 as conducting the same analyses as above along industry lines reveals only a single significant difference between them: sustainability reports from the semiconductor industry contain significantly ($p = 0.030$) fewer passive structures per sentence than the mining industry. As we see very little other evidence for a systematic difference in the language used between the semiconductor and mining industry and the effect size is small, this significance likely spawned from random chance. We will remain vigilant of salient differences between these industries in future research, but currently find insufficient justification for asserting any impact of industry on the language used in corporate sustainability reporting.

6. Conclusion and Future Research

KPMG's plea for more qualitative sustainability reporting rings as true in this study as in the few that have preceded it: traditional readability formulae place it amongst the most complex genres of writing. High-quality reporting means ensuring accessibility, and few companies will want to make the intense investment that exhaustive triple bottom line reporting requires only to frustrate their reports' readers. Sustainability-themed content readability contrasts unfavourably to even financially themed content. That means (sections of) these reports will likely tax the average reader of the latter, let alone the "average customer" that Farewell et al. mention, or many other sections of the sustainability report's far wider audience. Many companies claim to take a stakeholder-inclusive CSR stance, but a less specialised audience and an increase in textual complexity for relevant content make an inefficient combination at best.

This study also discovered variation in textual complexity between different varieties of English. The Flesch Reading Ease Score, Flesch-Kincaid Grade Level and Gunning Fog Index all indicated a marked increase in complexity for the Australian reports compared to the other varieties. However, these formulae only offer a limited view of the genre's complexity. Analysing a selection of 'deeper-level' syntactic features provided additional evidence for that trend, but also demonstrated that these reports may differ significantly in their use of language and its complexity beyond what the formulae can measure, such as is the case with the American reports' markedly more active structures or British reports' lower lexical density. These findings underline the importance of language variety as a predictor of linguistic complexity in addition to current models, such as Leuz et al.'s (2003) clusters of legal enforcement. As these results also demonstrate the merit of applying linguistic frameworks and language technology to corporate reporting, the next step in this avenue of research is to further investigate, through deep-level syntactic analysis, in which other linguistic aspects these reports differ between varieties of English and whether that variation is consistent with more general language models for the different Englishes. Experimental research might then extend such a comprehensive study with the aim of determining whether this linguistic variation in different language varieties also impacts cross-cultural stakeholders' appreciation of company performance.

It is not yet clear what determines linguistic complexity in sustainability reporting, especially given the observed increase in formal complexity for sustainability content over financial content. Similar studies into the readability of annual reports have not yet reached a consensus on whether company performance and annual report readability are correlated, i.e. whether companies attempt to conceal or obfuscate (deliberately or otherwise) unfavourable results. A crucial difference between sustainability reporting and financial reporting is that the former encompasses more areas of performance which might influence readability or incentivise concealment, which lends further salience to the observed readability gap. Poor environmental or social responsibility or corporate governance might lead a company to conceal negative outcomes just as weak financial performance might. In future studies, we will investigate how financial and non-financial performance variables impact the readability of reporting with respect to both sustainability reports and both varieties of CEO letters. Based on such an inquiry's result and the sustainability report's noted complexity, the genre require additional legislation to counteract risks of obfuscation that Cho et al. note (2012a, b).

The concept of sustainability will only become more deeply ingrained in many aspects of society. With over half of companies reporting on their CSR initiatives, the corporate sustainability report has become one of the primary means for those actors whose interests may not always appear to align with society's to maintain their social license by demonstrating clearly and transparently that their means of fulfilling their needs does not impact future generations' ability to meet theirs. Without

that clarity present in the language, however, companies might make these reports available to a large audience, but both corporate and scholarly initiatives such as KPMG’s reporting survey and this study may still impel them further towards transparent and accessible to that same audience.

ⁱ Tap Oil’s (2013) annual report for financial 2012, for example, has a single-page ‘Health, Safety, Environment & Community’ section out of a 105-page report.

ⁱⁱ Based on a 100.000-word corpus of British and American conversation, Precht finds that the former uses more modal verbs and the latter more emotive affect and emphatics.

ⁱⁱⁱ This incident, also commonly called the ‘Marikana Massacre’ in media (e.g. The Guardian 2012), occurred in South Africa in 2012 when protests over worker remuneration at a Lonmin-operated mine turned violent and police opened fire on protestors. This national tragedy made international headlines and raised substantial questions about the sustainability of Lonmin’s operations (ibid., Tolsi 2015).

^{iv}For example, Parsons & McKenna (2005), in a case study, points out that many statements such as “we set out to build enduring relationships with our neighbours characterised by mutual respect [...]”, offer no specific action that the company intends to take, nor a timeframe, nor, due to its vagueness, the ability to challenge this claim.

^v This more optimistic attitude on the readers’ part might be somewhat naive, given the aforementioned critical stance amongst scholars and the less-than-stellar results for readability that the analyses in Section 4 reveal.

^{vi} In order to prepare the data for processing, we stripped the texts of all paragraphs that were not formatted as running text. This removed, for instance, headings and subheadings, but also all numeric or non-full-sentence information contained in graphs and tables, dramatically reducing the length of many texts.

^{vii} We equate with one another, for the purposes of this study, the ‘CEO letter’, ‘Chairman’s Address’, and other variations thereupon that offer top-level management commentary on company performance.

^{viii} Thomson Reuters (2015) describes ASSET4 as a database which “provides objective, relevant and systematic environmental, social and governance (ESG) information based on 250+ key performance indicators (KPIs) and 750+ individual data points along with their original data sources.”

^{ix} In some cases, such as companies issuing biennial sustainability reports, we selected the report containing the greatest possible part of the calendar year 2012.

^x The full regular expression we used (via PowerGREP) was:

```
(?<=^d{0,3}\W*\d{0,3}\W*)(?!W*\tW*\tW*)([“” ’ ?]\w[^\t]+?(?<=[^\t]+? [^\t]+? [^\t]+?);:?!( and | or)?,?( “[”” ”,))?(?=[0-9]|\W)? ?$)
```

^{xi} We note that many of these formulae were initially designed to measure the level of reading material for teaching purposes, and are thus likely to become less accurate, especially in grade level estimations, on the extreme ends of the scale, as might well be the case for the Fog Index results in particular.

7. References

- Abu Bakar, A. S., & Ameer, R. (2011). Readability of Corporate Social Responsibility communication in Malaysia. *Corporate Social Responsibility and Environmental Management*, 18(1), 50–60. <http://doi.org/10.1002/csr.240>
- Adidas Group. (2013). Sustainability Progress Report 2012: Performance Counts. Retrieved from http://www.adidas-group.com/media/filer_public/2013/08/13/adidas_spr2012_full.pdf
- Bafilemba, F., Mueller, T., & Lezhnev, S. (2014). The Impact of Dodd-Frank and Conflict Minerals Reforms on Eastern Congo ' s Conflict.
- BBC. (2013). Bangladesh factory collapse toll passes 1,000. Retrieved from bbc.com/news/world-asia-22476774
- Beaman, K. (1984). Coordination and subordination revisited: Syntactic complexity in spoken and written narrative discourse. *Coherence in Spoken and Written Discourse*, 12, 45–80.
- Bebbington, J., Larrinaga-González, C., & Moneva-Abadía, J. M. (2008). Corporate social reporting and reputation risk management. *Accounting, Auditing & Accountability Journal*, 21(3), 337–361. <http://doi.org/10.1108/09513570810863932>
- Bogert, J. (1985). In Defense of the Fog Index. *Business Communication Quarterly*, 48(2), 9–12. <http://doi.org/10.1177/108056998504800203>
- Boiral, O. (2013). Sustainability reports as simulacra? A counter-account of A and A+ GRI reports. *Accounting, Auditing & Accountability Journal*, 26(7), 1036–1071. <http://doi.org/10.1108/AAAJ-04-2012-00998>
- Carnevale, C., & Mazzuca, M. (2014). Sustainability report and bank valuation: Evidence from European stock markets. *Business Ethics*, 23(1), 69–90. <http://doi.org/10.1111/beer.12038>
- Cho, C. H., Michelon, G., & Patten, D. M. (2012a). Enhancement and obfuscation through the use of graphs in sustainability reports: An international comparison. *Sustainability Accounting, Management and Policy Journal*.
- Cho, C. H., Michelon, G., & Patten, D. M. (2012b). Impression management in sustainability reports: An empirical investigation of the use of graphs. *Accounting and the Public Interest*, 12(1), 16–37. <http://doi.org/10.2308/apin-10249>
- 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. <http://doi.org/10.1016/j.aos.2006.09.009>
- Clatworthy, M. A., & Jones, M. J. (2003). Financial reporting of good and bad news: evidence from accounting narratives. (April 2015), 37–41. <http://doi.org/10.1080/00014788.2003.9729645>
- Coleman, E. B. (1964). The comprehensibility of several grammatical transformations. *Journal of Applied Psychology*, 48(3), 186–190. <http://doi.org/10.1037/h0040440>
- Colesanti, J. S. (2012). Demanding Substance or Form ? the Sec ' S Plain English Handbook As a Basis for Securities Violations.
- Courtis, J. K. (1995). Readability of annual reports: Western versus Asian evidence. *Accounting, Auditing & Accountability Journal*, 8(2), 4–17. <http://doi.org/10.1108/09513579510086795>

- Courtis, J. K. (1998). Annual report readability variability: tests of the obfuscation hypothesis. *Accounting, Auditing & Accountability Journal*. <http://doi.org/10.1108/09513579810231457>
- CPA Australia. (2013). *Sustainability Reporting: Practices, performance and potential*. Retrieved from <http://www.cpaaustralia.com.au/~media/corporate/allfiles/document/professional-resources/sustainability/sustainability-reporting-practice-performance-potential.pdf>
- Creese, A. (1991). Speech act variation in British and American English. *PENN Working Papers*, 7(2), 37–58.
- Dahlsrud, A. (2008). How corporate social responsibility is defined: An analysis of 37 definitions. *Corporate Social Responsibility and Environmental Management*, 15(1), 1–13. <http://doi.org/10.1002/csr.132>
- Dale, E., & Chall, J. S. (1948). A Formula for Predicting Readability: Instructions. *Educational Research Bulletin*, 27(2), 37–54 CR – Copyright © 1948 Taylor & Francis. <http://doi.org/10.2307/1473669>
- Deegan, C., Rankin, M., & Tobin, J. (2002). An examination of the corporate social and environmental disclosures of BHP from 1983-1997: A test of legitimacy theory. *Accounting, Auditing & Accountability Journal*. <http://doi.org/10.1108/09513570210435861>
- Dell', Orletta, F., Wieling, M., Cimino, A., Venturi, G., Dell'Orletta, F., & Montemagni, S. (2014). Assessing the Readability of Sentences: Which Corpora and Features? *Proceedings of the Ninth Workshop on Innovative Use of NLP for Building Educational Applications*, 163–173. Retrieved from <http://www.aclweb.org/anthology/W/W14/W14-1820>
- DuBay, W. (2004). *The principles of readability*. Costa Mesa: Impact Information. <http://doi.org/10.1.1.91.4042>
- Elkington, J. (1997). *Cannibals with forks. The triple bottom line of 21st century*. Retrieved from http://pdf-release.net/external/242064/pdf-release-dot-net-148_en.pdf
- Farewell, S., Fisher, I., & Daily, C. (2014). The Lexical Footprint of Sustainability Reports: A Pilot Study of Readability. In *American Accounting Association Annual Meeting and Conference on Teaching and Learning in Accounting*.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233. <http://doi.org/10.1037/h0057532>
- Flesch, R. (1949). *The Art Of Readable Writing*. (A. J. Gould, Ed.). New York and Evanston: Harper & Row.
- Global Reporting Initiative. (2013). *An Introduction to G4. The next generation of sustainability reporting*. Retrieved from <https://www.globalreporting.org/resourcelibrary/GRI-An-introduction-to-G4.pdf>
- Gunning, R. (1952). *The Technique of Clear Writing*. McGraw-Hill.
- 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*. Elsevier Ltd. <http://doi.org/10.1016/j.jclepro.2013.07.005>
- Halliday, M. A. K. (1989). *Spoken and written language. Language education*. Oxford: Oxford University Press.

- Harrison, S., & Bakker, P. (1998). Two new readability predictors for the professional writer: pilot trials. *Journal of Research in Reading*, 21(2), 121–138. <http://doi.org/10.1111/1467-9817.00049>
- Hrasky, S. (2012). Visual disclosure strategies adopted by more and less sustainability-driven companies. *Accounting Forum*, 36(3), 154–165. <http://doi.org/10.1016/j.accfor.2012.02.001>
- Infineon. (2013). *The Determining Factor. Infineon Technologies AG Annual Report 2012*. Retrieved from <http://www.infineon.com/dgdl/Infineon-GB2012-E.pdf?fileId=db3a30433b92f0e8013b989bf5cd15f3>
- Jones, M. J., & Shoemaker, P. A. (1994). Accounting narratives: A review of empirical studies of content and readability. *Journal of Accounting Literature*, 13, 142. Retrieved from <http://search.proquest.com.proxy.consortiumlibrary.org/docview/216304635>
- Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. *Technical Training, Research B*(February), 49. <http://doi.org/ERIC #:ED108134>
- Klare, G. R. (1963). *The Measurement of Readability*. Annes, Iowa: Iowa State University Press.
- KPMG. (2013). *The KPMG Survey of Corporate Responsibility Reporting 2013*. Retrieved from <http://www.kpmg.com/Global/en/IssuesAndInsights/ArticlesPublications/corporate-responsibility/Documents/kpmg-survey-of-corporate-responsibility-reporting-2013.pdf>
- Kumar, G. (2014). *Determinants of Readability of Financial Reports of U.S.-Listed Asian Companies. Asian Journal of Finance & Accounting* (Vol. 6). Retrieved from <http://www.macrothink.org/journal/index.php/ajfa/article/view/5695>
- Lehavy, R., Li, F., & Merkley, K. (2011). The Effect of Annual Report Readability on Analyst Following and the Properties of Their Earnings Forecasts. *The Accounting Review*, 86(3), 1087–1115. <http://doi.org/10.2308/accr.00000043>
- Leuz, C., Nanda, D., & Wysocki, P. D. (2003). Earnings management and investor protection: An international comparison. *Journal of Financial Economics*, 69(3), 505–527. [http://doi.org/10.1016/S0304-405X\(03\)00121-1](http://doi.org/10.1016/S0304-405X(03)00121-1)
- Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, 45(2-3), 221–247. <http://doi.org/10.1016/j.jacceco.2008.02.003>
- Lonmin. (2013). *Sustainable Development Report for the Year Ended 30 September 2012*. Retrieved from <http://sd-report.lonmin.com/2012/downloads/lonmin-online-sustainable-development-report-2012.pdf>
- Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 55–60). Retrieved from <http://nlp.stanford.edu/pubs/StanfordCoreNlp2014.pdf>
- McLaughlin, G. H. (1969). SMOG Grading - a New Readability Formula. *Journal of Reading*, 12(8), 639–646. Retrieved from http://harrymclaughlin.com/SMOG_Readability_Formula_G._Harry_McLaughlin_%281969%29.pdf

- Nearmy, T. (2014, March 27). New ASX guidelines to force sustainability reporting. *The Conversation*. Retrieved from <http://theconversation.com/new-asx-guidelines-to-force-sustainability-reporting-24885>
- Parsons, R., & McKenna, B. J. (2005). Constructing Social Responsibility in Mining Company Reports. In T. Lê & M. Short (Eds.), *Proceedings of the International Conference on Critical Discourse Analysis Theory into Research* (pp. 595–608). Retrieved from [http://195.130.87.21:8080/dspace/bitstream/123456789/262/1/Parsons & McKenna-constructing social responsibility.pdf](http://195.130.87.21:8080/dspace/bitstream/123456789/262/1/Parsons%20&%20McKenna-constructing%20social%20responsibility.pdf)
- Precht, K. (2003a). Great vs. lovely: Grammatical and lexical stance differences in American and British English. In C. Meyer & P. Leistyna (Eds.), *Corpus Analysis: Language Structure and Language Use* (pp. 133–151). Amsterdam: Rodopi.
- Precht, K. (2003b). Stance moods in spoken English: Evidentiality and affect in British and American conversation. *Text - Interdisciplinary Journal for the Study of Discourse*, 23(2). <http://doi.org/10.1515/text.2003.010>
- Rutherford, B. a. (2003). Obfuscation, textual complexity and the role of regulated narrative accounting disclosure in corporate governance. *Journal of Management and Governance*, 7(2), 187–210. <http://doi.org/10.1023/A:1023647615279>
- Sacconi, L. (2004). Corporate social responsibility (CSR) as a model of “ extended ” corporate governance . an explanation based on the economic theories of social contract , conformism. *Liuc Papers*, 147(Wieland 2003), 1–49.
- Securities and Exchange Commission. (1998). *A Plain English Handbook. How to create clear SEC disclosure documents*. Retrieved from <https://www.sec.gov/pdf/handbook.pdf>
- Smith, D. (2012, September 7). Marikana mine shootings revive bitter days of Soweto and Sharpeville. *The Guardian*. Retrieved from <http://www.theguardian.com/world/2012/sep/07/marikana-mine-shootings-revive-soweto>
- Smith, M., & Taffler, R. (1992). The Chairman’s Statement and Corporate Financial Performance. *Accounting & Finance*, 32(2), 75–90. <http://doi.org/10.1111/j.1467-629X.1992.tb00187.x>
- Stanford NLP Group. (2015). Stanford Parser. Retrieved from <http://nlp.stanford.edu:8080/parser/>
- Stanton, P., & Stanton, J. (2002). Corporate annual reports: research perspectives used. *Accounting, Auditing & Accountability Journal*. <http://doi.org/10.1108/09513570210440568>
- Story, J., & Neves, P. (2015). When corporate social responsibility (CSR) increases performance: exploring the role of intrinsic and extrinsic CSR attribution. *Business Ethics: A European Review*, 24(2), 111–124. <http://doi.org/10.1111/beer.12084>
- Tap Oil. (2013). *Annual Report 2012*. Retrieved from http://www.google.be/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0CCAQFjAA&url=http://www.tapoil.com.au/irm/Company/ShowPage.aspx/PDFs/1717-51411661/2012AnnualReport&ei=em-KVZ_fJobYyAODj6uAAg&usg=AFQjCNHgr9pL3gYW5wpt7Qde
- Temouri, Y., & Jones, C. (2014). Introduction: International business and institutions after the financial crisis. *Academy of International Business (uki)*, 21, 1–6. <http://doi.org/10.1057/9781137367204>

- Thomson Reuters. (2013). ASSET4 ESG Data Glossary. Retrieved from http://extranet.datastream.com/data/ASSET4%20ESG/documents/ASSET4_ESG_DATA_GLOSSARY_april2013.xlsx
- Thomson Reuters. (2015). Asset4. Retrieved from <http://www.trcri.com/index.php?page=asset4>
- Tilling, M. V. (2004). Refinements to Legitimacy Theory in Social and Environmental Accounting. *Commerce Research Paper Series, 06(04)*, 1–11.
- Tolsi, N. (2015). Marikana. Retrieved from <http://marikana.mg.co.za/>
- Total. (2013). *Working Together for Responsible Energy*. Retrieved from <http://www.total.com/sites/default/files/atoms/file/total-society-and-environment-report-2012>
- Townsend, S., Bartels, W., & Renaut, J.-P. (2010). Reporting Change. *Change*, 1–33. Retrieved from <http://www.sustainability.com/library>
- Van Marrewijk, M. (2003). Concepts and definitions of CSR and corporate sustainability: between agency and communion. *Journal of Business Ethics, 44*, 95–105. <http://doi.org/10.2307/25075020>
- World Commission on Environment and Development. (1987). *Report of the World Commission on Environment and Development: Our Common Future (The Brundtland Report)*. *Medicine, Conflict and Survival* (Vol. 4).

8. Tables

Table 1. Readability Formulae

Flesch Reading Ease Score	$206.835 - (1.015 * \text{average sentence length}) - (84.6 * \text{average syllables per word})$
Flesch-Kincaid Grade Level	$(0.39 * \text{average sentence length}) + (11.8 * \text{average syllables per word}) - 15.59$
Gunning Fog Index	$0.4 * (\text{average sentence length} + \text{percentage of polysyllabic words})$

Table 2. Corpus Composition

Row Labels	Mining	Oil	Semiconductors	Apparel	Grand Total
Fin. oriented CEO Letter(s)	95	82	30	12	219
USA	11	35	22	4	72
UK	18	11	2	0	31
Europe	17	15	5	8	45
Australia	44	18	1	0	63
India	5	3	0	0	8
Sust. Oriented CEO Letter(s)	38	35	12	3	88
USA	4	14	8	2	28
UK	14	7	1	0	22
Europe	9	8	3	1	21
Australia	8	5	0	0	13
India	3	1	0	0	4
Sustainability Report	78	59	16	10	163
USA	9	18	10	2	39
UK	18	11	2	0	31
Europe	17	16	4	8	45
Australia	29	11	0	0	40
India	5	3	0	0	8
Totals					
USA Count	24	67	40	8	139
UK Count	50	29	5	0	84
Europe Count	43	39	12	17	111
Australia Count	81	34	1	0	116
India Count	13	7	0	0	20
Grand Total	211	176	58	25	470

Table 3. Analysis for sustainability reports only

Sustainability Reports Only	Flesch Score		Kincaid Score		Fog Score		Lexical Density		Subordinators / Sentence		Parse Tree Depth		Passives / Sentence	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Regions														
Australia	9.68	10.91	18.35	3.21	23.07	3.68	0.64	0.02	0.53	0.37	10.91	1.43	0.35	0.13
Europe	19.03	5.57	16.46	1.54	20.80	1.69	0.63	0.02	0.43	0.11	10.35	0.84	0.30	0.08
India	15.94	2.90	16.46	0.62	20.86	0.74	0.66	0.02	0.27	0.09	9.87	0.50	0.30	0.06
UK	17.96	5.02	16.50	1.07	21.04	1.24	0.63	0.01	0.48	0.09	10.55	0.65	0.30	0.08
USA	17.20	6.42	16.40	1.40	20.79	1.41	0.65	0.02	0.49	0.18	10.17	0.69	0.22	0.11
Industries														
Apparel	20.53	6.15	16.43	1.68	20.58	1.90	0.62	0.02	0.47	0.14	10.59	1.00	0.26	0.08
Mining	14.98	8.73	17.00	2.08	21.55	2.38	0.64	0.02	0.47	0.27	10.46	1.09	0.32	0.12
Oil	16.19	8.07	17.03	2.46	21.56	2.71	0.64	0.02	0.48	0.19	10.47	0.91	0.28	0.11
Semiconductors	16.89	4.37	16.39	0.89	20.64	0.90	0.64	0.02	0.44	0.10	10.29	0.77	0.23	0.08
Standalone Sustainability Report														
Yes	17.17	4.74	16.50	1.07	20.89	1.23	0.64	0.02	0.44	0.10	10.32	0.61	0.27	0.08
No	14.54	10.56	17.39	2.82	22.00	3.15	0.63	0.03	0.50	0.31	10.62	1.28	0.32	0.14
Means	15.94	8.08	16.92	2.12	21.40	2.39	0.64	0.02	0.47	0.22	10.46	0.99	0.29	0.11

Table 4. Analysis for full corpus

Full Corpus	Flesch Score		Kincaid Score		Fog Score		Lexical Density		Subordinators / Sentence		Parse Tree Depth		Passives / Sentence	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Genres														
CEO Letter(s)	26.02	8.13	15.78	1.94	20.11	2.17	0.62	0.02	0.52	0.20	10.95	1.24	0.21	0.09
Sustainability Letter	19.80	9.29	16.37	2.18	20.65	2.41	0.60	0.03	0.62	0.27	11.01	1.38	0.18	0.10
Sustainability Report	15.94	8.08	16.92	2.12	21.40	2.39	0.64	0.02	0.47	0.22	10.46	0.99	0.29	0.11
Means	21.36	9.49	16.28	2.11	20.66	2.36	0.62	0.03	0.52	0.23	10.79	1.21	0.23	0.11