Corporate Social Responsibility in the IBEX 35: A Corpus-based Study of CSR Reports

Carla Pérez Almendros*, Luis Espinosa Ankeb, Rosa Torres Valdés
c

a. Asociación Educativa Integral del Raval, Rambla del Raval, 7bis bajos, Barcelona 08001, Spain
b. Universitat Autònoma de Barcelona, Campus de Bellaterra, Cerdanyola del Vallès, 08193, Spain
c. Departamento de Comunicación y Psicología Social, Universidad de Alicante, San Vicente del Raspeig, 03690, Spain

Abstract

We present a corpus-based study of the Corporate Social Responsibility Reports issued by companies who operate in the Spanish Stock Market IBEX 35. Natural Language Processing methods are used for extracting informative collocations and measuring the textual complexity of these texts. Based on our results, we claim that (1) these reports are in general simpler and more redundant than general language, (2) there is a tendency towards the increasing use of the term sustainability to describe SR actions, and (3) integrated reports seem to be written for shareholders rather than the general public.

1. Introduction

The term “Corporate Social Responsibility” (CSR) is generally used to describe a company’s commitment towards topics demanded by society (Méndez Picazo, 2005). It refers to a form of corporate management where they voluntarily attempt to give back to society and the environment part of what they take from them during the development of their activity (Frederick et al., 1992). According to Van Marrewijk (2003), this is in most cases carried out together with their publics.

* Corresponding author. Tel.: +34-685-449-364 .
E-mail address: carlaperezalmendros@gmail.com
CSR has had a positive impact among the general public (Alvarado Herrera and Walesska Schelesinger, 2008). However, it has also been criticized by those claiming that CSR can become an instrument for legitimating any company’s actions in front of its members or shareholders (Handelman and Arnold, 1999), or for improving its corporate image (Sen and Bhattacharya, 2001). While the concept Corporate Social Responsibility was formally coined in the 50s (Díaz Molina and Facal Santiago, 2011), the development in different areas such as socio-economic, environmental, political and corporate ethics, and so forth have contributed to blur this term’s semantic boundaries. On one hand, there is a plethora of concepts related with this activity, and on the other, terms are still being introduced for describing CSR actions, such as “corporate citizenship” (Carroll, 2000), “social responsiveness” or “social sensitivity” (Suñerán, 1995). Likewise, authors like Valor (2005) or Moneva (2005) have shown concerns on how the term CSR is used or its interpretation, respectively.

This interest for CSR crystalizes in pieces of information like the following: Between 1999 and 2005 the number of companies that had endorsed the United Nations Global Compact has increased from 38 to 2227, out of which 220 are Spanish (Nieto, 2005). Generally, the CSR developed by a company can be evaluated by means of a range of variables, one of them being the analysis of its yearly CSR report. Hence, we assume a direct relationship between the actions that a company carries out in the field of CSR and how these are deployed in its CSR report. This work presents a corpus-based study of the CSR reports published by Spanish companies who are present in the Spanish stock market IBEX 35 in the year 2011.

We present statistically motivated information extracted from the documents that make up our corpus with the objective of shedding light on the following research hypotheses.

1) The writing style of the CSR reports in the IBEX 35 context tends to be made according to fixed and recurrent lexical choices typical of this field, which might lead to repetitive and tedious texts.

2) There seems to exist a tendency of term-switching, where Corporate (Social) Responsibility leaves room for sustainability, which might reflect a close linking with production processes, social change or environmental care.

3) Integrated reports do not emphasize CSR or sustainability as much as the dedicated reports.

In order to confirm or reject these hypotheses, we follow three main approaches: Firstly, we extract and score collocations according to the metric log-likelihood ratio in order to unveil any semantically relevant pattern. Secondly, we carry out a frequency-based study of the terms Responsabilidad (Responsibility), Sostenibilidad (Sustainability), Responsabilidad Corporativa (Corporate Responsibility), Responsabilidad Social (Social Responsibility), Responsabilidad Social Corporativa (Corporate Social Responsibility), Responsabilidad Social Empresarial (Organizational Social Responsibility), Responsabilidad Empresarial (Organizational Responsibility), RSC (CSR) and RSE (OSR). Finally, we present data on the textual complexity of these reports, where factors like length, vocabulary, lexical density, lexical richness and lexical variety are taken into consideration.

2. Related Work

It is generally agreed that the relationship between Social Responsibility and Business started in 1953 with Bowen’s work Social Responsibilities of a Businessman (Saavedra Robledo et al., 2010). From there, a philosophical time where the aims of the company were debated followed. Precisely, in the field of communication and Public Relations, Lucien Magrat’s speech in 1955 at the Faculty of Law of Bourdeaux explicitly linked the company and its social responsibilities (Torres, 2008).

According with an analysis by Saavedra Robledo et al. (Ib.), the concept Corporate Social Performance is coined in 1970, as a way of measuring CSR. The concept Social Balance was later on coined in the 1980s as well as the interaction between CSR and sectorial development and corporate citizenship concepts.

The etimological origin of the term responsibility can be traced back to the Latin responsum, to answer. Additionally, let us highlight the fact that the concept of responsibility is “above all, moral and judicial” (Solano Fleta, 1995). And, thus, “responsibility can be defined as a person’s capacity to know and accept the consequences derived from their acts” (Ib.).

Within the theoretical and methodological framework of this work, ethical actions have been linked with the importance of the public opinion (Friedman, 1970; Orr, 1974).
A current perspective takes us to González (2005), who join growth, competitiveness and Corporate Social Responsibility, and furthermore point out, in the context of Spain, that a trivialization of the term leads to great confusion in relation to its meaning. This might even lead to misinterpreting CSR as a term for social action, which becomes the finest form of promotion of CSR policies, which might be over-reduced to donations to social causes.

Moreover, González (op cit.) suggests that CSR embodies environmental, social and work-related aspects, and thus comments that it is a term constantly evolving, and which should not be considered as a marketing tool.

The guidelines published by international companies which operate in this field relate in their statements CSR with the environment, social issues, labor or business ethics, and in all cases the focus seems to be put on increasing sustainability.

Taking into account the above review, we claim that Social Responsibility consists in responsible management of them. This management enables the company to voluntarily design and implement actions in different areas, seeking a balance of interests and expecting a legitimate minimum return in the quality of its internal and external relations, in short or mid-term, and without failing to comply with ethical or legal regulations (Torres, 2008).

3. Method

3.1. Collocations

We take as a starting point Firth’s classic “you shall know a word by the company it keeps” (1957:11), and theorize that those concepts that occur frequently might reveal the idiosyncrasy of the text and, consequently, the issuing company and the target audience. Ranking is performed using the log-likelihood ratio association measure. The method consists in extracting bigrams and trigrams and score them according to this metric. The 10 highest scoring collocations are presented. For this task, stopwords have been removed.

3.2. Likelihood-ratio

Following Manning & Schütze (1999:172,173), likelihood ratio is an approach that can be used to assess to what extent the fact that a word $n$ follows a word $m$ indicates that we are in front of a collocation, or if it is pure chance. This statistical measure is, in addition, appropriate for rare lexical combinations, those who do not occur much in a text despite being potentially informative. See Dunning (1993) and Manning & Schütze (1999:151-176) for the mathematical formalization of this metric.

3.3. Text complexity

We perform a comparative analysis of the textual complexity of the CSR reports, as opposed to the reference corpus Ancora (Taulé et al., 2008), which we exploit as an example of general language. In this way, we aim at confirming or rejecting the hypothesis that CSR reports issued by companies operating in the IBEX 35 stock market are in general redundant, show less lexical variation than standard Spanish, and that this reflects a stiff narrative which is highly dependent of the field and the audience involved in the communicative process. Each report is scored according to:

Type-token ratio:

Type/token ratio ($ttr$) is the result of dividing the total number of different words by the total number of occurrences, multiplied by 100. The higher the result, higher is the lexical diversity of the sample. However, a low result indicates a higher repetition. This means that less knowledge of the language is required to fully understand the text. The more a text grows, the probability to encounter new words decreases gradually.
Lexical diversity:
We understand by lexical diversity the number of times each word is repeated on average, as it appears in Bird et al. (2009). It is simply a metric that inverts the numerator and denominator of the ttr.

Lexical density:
Lexical density is the proportion of semantically-rich words (nouns, verbs, adverbs, adjectives) in relation with words which carry no meaning (prepositions, conjunctions, auxiliaries, etc.).

Guiraud’s Index:
Since ttr is highly sensitive to the text length, alternatives have been proposed. One of the best known is the Guiraud’s index (Guiraud, 1954), which simply adds a calculation of the square root of the ttr’s denominator. It is claimed that by doing this, sensitivity to text length is reduced.

4. Results and Conclusions

4.1. Results

After cleaning the corpus thoroughly, and using the methods described in Section 3, we develop frequency lists of the most popular terms used to refer to this organizational philosophy. We study the following keywords: Responsabilidad (Responsibility), Sostenibilidad (Sustainability), Responsabilidad Corporativa (Corporate Responsibility), Responsabilidad Social (Social Responsibility), Responsabilidad Social Corporativa (Corporate Social Responsibility), Responsabilidad Social Empresarial (Organizational Social Responsibility), Responsabilidad Empresarial (Organizational Responsibility), RSC (CSR) and RSE (OSR). The two acronyms are clearly underused in our corpus. We present figures for the rest in Table 1.

<table>
<thead>
<tr>
<th>Term</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responsabilidad (Responsibility)</td>
<td>1687</td>
</tr>
<tr>
<td>Sostenibilidad (Sustainability)</td>
<td>1510</td>
</tr>
<tr>
<td>Responsabilidad Corporativa (Corporate Responsibility)</td>
<td>444</td>
</tr>
<tr>
<td>Responsabilidad Social (Social Responsibility)</td>
<td>392</td>
</tr>
<tr>
<td>Responsabilidad Social Corporativa (Corporate Social Responsibility)</td>
<td>160</td>
</tr>
<tr>
<td>Responsabilidad Social Empresarial (Organizational Social Responsibility)</td>
<td>8</td>
</tr>
<tr>
<td>Responsabilidad Empresarial (Organizational Responsibility)</td>
<td>7</td>
</tr>
</tbody>
</table>

The concept Social Responsibility is much more frequent than the classic Corporate Social Responsibility, which confirms that companies are interested in other topics in addition to social issues.

While the term responsibility is still the most used, sustainability is also a very frequent term, which might suggest a tendency towards a change in the preferred term if we consider that the latter was introduced in the late 80s and is already being used thoroughly.
Table 2. Length, vocabulary and lexical density, richness and variation.

<table>
<thead>
<tr>
<th>Report</th>
<th>Corpus length</th>
<th>Corpus vocab.</th>
<th>Lexical density</th>
<th>Lexical richness</th>
<th>Type Token Ratio</th>
<th>Guiraud Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abengoa</td>
<td>57517</td>
<td>6214</td>
<td>60.23</td>
<td>9.25</td>
<td>0.10</td>
<td>25.91</td>
</tr>
<tr>
<td>Acerinox</td>
<td>28020</td>
<td>4411</td>
<td>64.09</td>
<td>6.30</td>
<td>0.15</td>
<td>26.53</td>
</tr>
<tr>
<td>ACS</td>
<td>36953</td>
<td>4407</td>
<td>59.82</td>
<td>8.38</td>
<td>0.11</td>
<td>22.92</td>
</tr>
<tr>
<td>CaixaBank (Integrated)</td>
<td>69373</td>
<td>6011</td>
<td>61.02</td>
<td>11.54</td>
<td>0.08</td>
<td>22.82</td>
</tr>
<tr>
<td>Inditex (Integrated)</td>
<td>139428</td>
<td>8744</td>
<td>60.93</td>
<td>15.94</td>
<td>0.06</td>
<td>23.41</td>
</tr>
<tr>
<td>Indra (Integrated)</td>
<td>70941</td>
<td>6821</td>
<td>61.36</td>
<td>10.40</td>
<td>0.09</td>
<td>25.60</td>
</tr>
<tr>
<td>Reference corpus</td>
<td>237117</td>
<td>26545</td>
<td>66.76</td>
<td>8.93</td>
<td>0.11</td>
<td>54.51</td>
</tr>
</tbody>
</table>

Some of the reports analyzed are presented for illustrative purposes in Table 2. We can see dedicated reports as well as integrated reports and information about their lexical distribution. The AnCora-Es corpus is more than 500k words long and is essentially a collection of newswire texts.

Integrated reports are in all cases longer (some of them even four times longer) than the average dedicated report. Regarding lexical density, there seems to be a fair balance between these two types of reports. Lexical variation results describe the degree of complexity of a text according to its score. In general, we can see that the CSR reports, whether these are integrated or dedicated, present a similar index among them, and interestingly, a much lower score than the reference corpus.

Tables 3 and 4 present the most frequent collocations found in the corpus, the first focused on dedicated reports, and the second on integrated reports.

Table 3. Highest occurring collocations in the CSR reports.

<table>
<thead>
<tr>
<th>Report</th>
<th>Collocations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abengoa</td>
<td>(sustainable development) desarrollo sostenible; (environmental services) servicios medioambientales; (greenhouse effect) efecto invernadero; (social responsibility) responsabilidad social; (corporate governance) gobierno corporativo; (our team) nuestro equipo; (climate change) cambio climático; (human rights) derechos humanos.</td>
</tr>
<tr>
<td>Banco Popular</td>
<td>(human rights) derechos humanos; (social responsibility) responsabilidad social; (environment) medio ambiente; (human resources) recursos humanos.</td>
</tr>
<tr>
<td>Bolsas y Mercados Españoles</td>
<td>(the employees) los empleados; (2011 exercise) ejercicio 2011; (the markets) los mercados; (social responsibility) responsabilidad social; (collective agreement) convenio colectivo; (meeting) junta general; (environment) medio ambiente; (global compact) pacto mundial; (financial markets) mercados financieros; (united nations) naciones unidas.</td>
</tr>
<tr>
<td>Ferrovial</td>
<td>(environment) medio ambiente; (corporate responsibility) responsabilidad corporativa; (energy efficiency) eficiencia energética; (good practice) buenas prácticas; (global compact) pacto mundial; (the companies) las empresas; (the employees) los empleados; (collaboration with) colaboración con; (human rights) derechos humanos.</td>
</tr>
<tr>
<td>Repsol</td>
<td>(human rights) derechos humanos; (environment) medio ambiente; (corporate responsibility) responsabilidad corporativa; (energy efficiency) eficiencia energética; (the employees) los empleados; (greenhouse effect) efecto invernadero; (the people) las personas; (the emissions) las emisiones; (the rights) los derechos; (indigenous communities) comunidades indígenas.</td>
</tr>
</tbody>
</table>
Table 4. Highest occurring collocations in integrated reports.

<table>
<thead>
<tr>
<th>Report</th>
<th>Collocations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enagás (Integrated)</td>
<td>(annual accounts) cuentas anuales; (corporate governance) gobierno corporativo; (meeting) junta general; (2011 exercise) ejercicio 2011; (environment) medio ambiente; (internal auditing) control interno; (the advisors) los consejeros; (bylaws) estatutos sociales; (social capital) capital social; (financial information) información financiera; (senior management) alta dirección; (relation with) relación con; (corporate name) denominación social.</td>
</tr>
<tr>
<td>BBVA (Integrated)</td>
<td>(bbva group) grupo bbva; (united states) estados unidos; (fair value) valor razonable; (of the group) del grupo; (consolidated balance) balances consolidados; (financial services) servicios financieros; (2011 exercise) ejercicio 2011; (representative values) valores representativos; (balance) los balances; (financial assets) activos financieros; (net equity) patrimonio neto; (the accounts) las cuentas;</td>
</tr>
<tr>
<td>CaixaBank (Integrated)</td>
<td>(global compact) pacto mundial; (social work) obra social; (of the risk) del riesgo; (united nations) naciones unidas; (the clients) los clientes; (strategic plan) plan estratégico; (corporate government) gobierno corporativo; (the companies) las empresas; (the employees) los empleados</td>
</tr>
<tr>
<td>Inditex (Integrated)</td>
<td>(meeting) junta general; (of the council) del consejo; (financial information) información financiera; (yearly accounts) cuentas anuales; (2011 exercise) ejercicio 2011; (corporate governance) gobierno corporativo; (managing director) consejero delegado; (the advisors) los consejeros; (bylaws) estatutos sociales; (financial states) estados financieros; (the statutes) los estatutos;</td>
</tr>
<tr>
<td>Indra (Integrated)</td>
<td>(ethical code) código ético; (environment) medio ambiente; (human rights) derechos humanos; (internal audit) auditoría interna; (corporate responsibility) responsabilidad corporativa; (of the council) del consejo; (meeting) junta general; (with disabilities) con discapacidad; (profesional behaviour) conducta profesional; (the systems) los sistemas; (air traffic) tráfico aéreo; (global compact) pacto mundial; (accessible technologies) tecnologías accesibles.</td>
</tr>
</tbody>
</table>

Both tables show that the most frequent collocations in each type of report are, in addition, recurrent across most texts. This also suggests that lexical variation is scarce due to similar lexical choices. The bigrams (environment) medio ambiente, (human rights) derechos humanos and (social/corporate responsibility) responsabilidad social/corporativa are the most frequent. However, in integrated reports these are not concepts that occur as frequently as others, which are more related to the field of finance and business than social work or sustainability.

4.2. Conclusions

In the light of these results, the following are conclusions derived from an interpretation of these results.

As we have observed in the use of terms used to describe this organizational philosophy, social responsibility is still a widely extended concept, in spite of it being a matter of discussion for limiting the scope of the responsibility to the company or corporation. In addition, the item social is frequently dropped from the traditional term (CSR), which implies that companies are aware of their responsibility not being only social, but also affecting other policy fields. While responsibility is still the widest used term, sustainability is gaining popularity as its position in the frequency table suggests. In fact, some reports like the ones by Acciona, Banco Santander or Iberdrola, are entitled (corporate sustainability report) memoria de sostenibilidad de la empresa. Likewise, in the case of Telefónica, the report is named (Sustainability and Corporate Responsibility Report) Informe de Sostenibilidad y Responsabilidad Corporativa. In this same direction, Mediaset’s report, although it has the title (Corporate Responsibility Report 2011) Informe de Responsabilidad Corporativa 2011, it explicitly mentions the (sustainability aspects dealt with in the report) los aspectos de sostenibilidad tratados en la memoria. Let us finally highlight the report by Repsol, where a vast number of examples of the use of “sustainability” can also be found with a meaning similar to “corporate responsibility”, as in (The objectives of the sustainability plan at closing date December 2011 are reported) Se informa de los objetivos del plan de sostenibilidad con fecha de cierre diciembre 2011.

Responsibility has more than one sense. It not only refers to that responsibility that the company consciously acknowledges to alleviate its effects in the environment. Sustainability, however, is a term that tends to be used interchangeably with the sense of responsibility that we are discussing. Since the latter was coined in the late 80s by Harlem, in the United Nation’s Report on our common future (Naredo, 2004), and popularized later on, its use seems to mark a turning point in its taking over in responsibility.

After reviewing the lexical composition of the CSR reports, the following conclusions are drawn:
This corpus features, in general terms, similar scores for vocabulary variation and lexical density among its texts, and much lower average scores than the reference corpus. In addition, lexical variation is in all cases much lower (less than 50%) than in our reference corpus. We claim that these results suggest the following:

1) CSR reports in the IBEX 35 companies can be seen as a standalone textual genre due to the great similarities existing among them, and the significant differences found with a reference corpus.

2) This textual genre presents texts with a reduced vocabulary, which makes them easy to understand to lay audiences, but as a tradeoff, these seem to be tedious and repetitive reads.

3) Lexical choices are almost the same in all the reports. The same topics are covered, and in most cases with the same terms, again, suggesting a very low lexical variation.

Having reviewed the lexical choices in integrated and dedicated CSR reports, and how lexical choices reflect the attitude of the company in each case, we conclude that this seems to be closely related to the potential target audience of these reports. While concepts related to human rights, environment or sustainable development are present in both cases, they are scarce in the integrated reports, where the focus seems to be more in financial or economic aspects. For this reason, it seems clear that these integrated reports are targeted towards the company’s shareholders. This has a major implication: An assumption that shareholders are more interested in the financial benefits of the company than its social actions.

5. Discourse Analysis

So far, we have presented corpus-based quantitative data derived from frequency distribution and statistical analyses. We hypothesize, however, that human assessment using Discourse Analysis methods would be recommended in order to reveal the meaning of the terms highlighted.

Language in the business field is highly contextual (Thomas, 1997), and depends, in this case, of the company's context, the field, the economic situation, the state of development in the area of CSR, and so forth. Therefore, we point towards a future line of research where a contrastive analysis of the terminology studied is carried out. We attempt to do this by looking at how these concepts are provided with meaning in our specialized corpus, on one hand, and in business-related or newswire texts, on the other.

Furthermore, we aim at exploring with greater detail the term Responsabilidad Social (Social Responsibility) in our Ibex-35 corpus as well as in texts published by international bodies that are considered a reference in this matter, such as the Millennium Development Goals or the World Compact, by United Nations, or The Equator Principles. In this way, elements like context or the interpersonal function of language (Halliday, 1985) can be incorporated as a factor that influences how this concept is treated.

Next, a human analysis of the discoursive treatment given to the most frequent collocations is expected to complete our research. We are aware of the fact that having a term or concept appearing repeatedly does not imply it is referring to more actions related to it. We hypothesize that Discourse Analysis techniques could unveil situations where a popular or altruistic term used repeatedly might in fact mean the contrary, an attempt to hide a lack of action by the issuing company in this respect. Let us examine a few of the sentences where the most frequent collocations are included. Specifically, we exemplify this with the term derechos humanos (human rights), which appears as a collocation of statistical significance, and its presence in Abengoa's report.

"Abengoa, from its international company position, advocates for sustainable growth based on the respect to human rights"

"showing awareness of the responsibility to ensure that human rights are respected, Abengoa assesses and reviews all its operations, measuring its impact in terms of human rights"

"during 2011 there have been no complaints related to human rights"

"training is one of the most powerful tools that Abengoa has to prevent incidents related to the violation of human rights from happening”

We can observe that, despite the inclusion of the term human rights in these sentences, there is not much being said about it. What does Abengoa specifically do for the defense of human rights? Despite it being left unexplained, the constant repetition of the term would convince the reader that Abengoa fights in favour of the human rights.
This could take us to a deeper question: What does Social Responsibility mean for these companies? Adapting to what international bodies require, or to what society demands? Do these corporations make conscious decisions in this respect, and act based on their corporate philosophy and values?

Let us look at the following passage from the report by Banco de Sabadell (Bank of Sabadell):
"the objective is to identify risks and opportunities in terms of corporate social responsibility to maximize the creation of value and meet the expectations of the interested parts".

This sentence seems to imply that the company is undertaking its Social Responsibility policy as a way of generating value towards its publics, and taking into consideration their inclinations and preferences. This claim seems to approach the definition of Social Responsibility made by Méndez Picazo (2005) or Van Marrweijk (2003), and seems to be coherent with the actions described along the text under analysis. This fact, which seems to be fairly infrequent across reports from other Ibex-35 companies, leads to the conclusion that Bank of Sabadell appears to be one of the companies that best have understood the essence of Social Responsibility and how it is reflected by the scientific community.

6. Future work

We have discussed a novel corpus-based approach to the analysis of CSR reports. While automatic text processing allows us to extract relatively complex linguistic patterns to understand and interpret a text, we are not alien to the need of qualitative analyses to provide more comprehensive and reliable results. We therefore propose an individualized assessment of each text incorporating Discourse Analysis methods to set the degree of correlation between the frequency of a term related to CSR and the actual actions taken by the company. We believe it is too speculative to claim that there is a direct correlation between the fact that, for example, human rights appears a lot in a report and that the company is in fact preoccupied and taking actions for improving human rights in the world.

On the other hand, we propose to extend our automatic approach to a comparative study between CSR reports and concepts associated to them, such as the Millennium Development Goals, the Equator Principles or the Global Compact. A corpus-based study will shed light over the textual similarity between a CSR report and these reference objects, hence providing figures on the relevance of the company’s activities in this context. Followed by Discourse Analysis methods, human and statistical data could be contrastively analyzed.

Finally, we leave an open door for incorporating specialists in Public Relations and CSR, who can perform an evaluation of the degree of complexity or interest that these reports generate, always aiming at confirming or rejecting empirical data obtained automatically.

Acknowledgements

This work has been partially funded by Obra Social “la Caixa”, Programa de Becas 2011 - 2013.

References


