Similarity of management reports of Brazilian firms

Ludmila de Melo Souza*

University of Brasilia,
Campus Darcy Ribeiro,
Prédio da FACE, Asa Norte,
70910900, Brasília, DF, Brazil
and
University of São Paulo,
Avenida Professor Luciano Gualberto,
908, Butantã, São Paulo/SP, 05508-010, Brazil
Email: ludmilamelo@usp.br
*Corresponding author

César Augusto Tibúrcio Silva

University of Brasilia,
Campus Darcy Ribeiro,
Prédio da FACE, Asa Norte,
70910900, Brasília, DF, Brazil
Email: cesartiburcio@unb.br

Abstract: Management reports (MR) are used as companies' communication tool for broadcasting a message or image to its shareholders. This study presents an investigation of the structure of MR with the purpose of verifying if there is a repetition of data. The aim of the study was to identify which variables explain the similarity between the MR that were published by companies listed in the Brazilian stock market. The results signalled that most of these companies had only specific modifications at their reports. The fact that other independent variables were not shown as explanatory at the obtained model may suggest that the maintenance of similar reports' structures is a usual practice, regardless of their characteristics.

Keywords: management reports; similarity; communication; Brazilian firms; disclosure; business narratives; Brazil.

Biographical notes: Ludmila de Melo Souza is an Assistant Professor at University of Brasilia (UnB). She received his Master's degree in Accounting and Bachelor of Accounting from the UnB and a doctoral student in Accounting and Controllership at the University of São Paulo (USP).

César Augusto Tibúrcio Silva is a Full Professor at University of Brasilia (UnB). He received his Bachelor in Business Administration from UnB, Bachelor of Accounting from Unieuro, Master degree in Business Administration from UnB and PhD in Accounting and Controllership from the University of São Paulo (USP).

1   Introduction

The management reports (MR) have been used by companies as a communication tool for broadcasting a message to its shareholders (Herremans and Ryans, 1995). The narrative communication through the MR is understood as critical for the achievement of the desired change in the quality of companies' reports (Beattie et al., 2004). Llewellyn (1999) argues that the narrative is presented as a way to explain the thinking and to persuade, and is as legitimate as those supported by quantitative reasoning. For Bryan (1997), narrative reports provide for information that cannot be easily included in the financial statements. Despite such narrative reports' alleged legitimacy those have been underutilised by the social sciences (Llewellyn, 1999). This is primarily due to the widespread notion that reports are biased. Dye (2001) argues that any entity that is able to perform a disclosure will present information which are favourable and omit those that are unfavourable.

We emphasise that there is no obligation to perform an independent audit on information contained in the MR; however, they may be optionally reviewed (Springer, 1992). Rutherford (2003) points out two problematic aspects regarding the regulation and standardisation of accounting narratives: the difficulty in establishing standards with certain accuracy and in the monitoring by the regulators. Whereas MR present understandable, relevant, reliable and comparable information, this study aims at checking if there is any duplication in information disclosed from one financial year to another. There is an understanding that companies tend to reproduce certain report structure throughout the years. This practice may reduce the investor's ability to make decisions based on financial accounts and/or mislead the same. Furthermore, this practice represents a lost opportunity to promote the company's achievements, in addition to suggesting certain neglect in disclosing information to the market.

The hypotheses of this study are as follows:

1   over time companies tend to reproduce the structures of its MR

2   there are no significant changes in the amount of information disclosed in the MR during the study period

3   according to their size, the companies tend to present roughly similar MR in subsequent periods

4   companies with greater political costs[1] tend to present less similar MR

5   companies with better performance indicators tend to present less similar MR

6   companies with more complex organisational structures tend to present less similar MR

7   companies with greater indicators of ownership concentration tend to present more similar MR

8   companies with higher risks of investments tend to present more similar MR.

This research is structured on four sections, and the introduction hereof. The literature review is presented below. The third section describes the research procedures. The fourth section presents the results found and analysis obtained. The last section addresses the final considerations, the research limitations as well as suggestions for future studies.

2   Background

2.1   Annual reports

The existing literature on annual reports takes several approaches to the analysis of the financial accounts. For Beattie et al. (2004), the study of MR is divided into two main methodologies: the subjective and the semi-objective methodology. An example of subjective methodology would be the 'analysts' rating'. In methodologies classified as semi-objective, the authors make a division between those using the disclosure index (using binary variables, weights, and hierarchical categories) and those guided by textual analysis (content analysis, readability, and linguistic analysis).

Several studies on the MR were developed have been developed internationally. Marston and Shrives (1991) performed a review on the use of the disclosure index in the accounting research, particularly in narrative annual reports. Jones and Shoemaker (1994) reviewed 32 legibility studies, 26 of them related to the use of Flesch readability in measuring the readability of MR. The researchers found that there is a general agreement that MR are, most often, classified as 'hard' or 'very hard' reading. Abrahamson and Amir (1996) performed a study aimed at verifying if information contained in the section message to shareholders of the MR were among the variables analysed by investors when analysing the accounting numbers. Courtis (1998) found that companies with MR presenting low readability disclose bad news. Sydserff and Weetman (1999) suggested that the textual content is a potential tool for analysing the financial accounts and for performing association tests. Boo and Simnett (2002) found that companies that undergo financial constraints and do not evidence this situation in their reports are more likely to face bankruptcy than those in difficulties but that show such fact in the MR, most often optimistically. Yuthas et al. (2002) found that those companies that use the MR for broadcasting reliable and trustworthy information do anticipate gains or losses. Zeller et al. (2012) have studied annual report design for 500 companies. They concluded that US companies have improved the quality and presentation of non-financial and financial information in annual reporting and suggested that fewer companies are using the annual report as an impression management vehicle. Van Geyt et al. (2014) explored if high-quality corporate communication contributes to reducing insider trading profitability and information asymmetry. The authors verified that has a negative association between corporate communication quality and insider trading profitability. Lee and Hong (2014) aimed to complement the manual extraction process by using text mining techniques because the great volume of the data present on the annual reports. Then, they found temporal changes in the business model concept over time. Dhanani and Connolly (2015) examined aspects of non-governmental organisational (NGO) annual reporting using

Habermas' theory of communicative action. The results indicated that organisational discloses are characterised by communicative action through the provision of truthful disclosures, in a manner to improve their understandability. Bambagiotti-Alberti et al. (2014) have studied the quality of disclosure and the level of accountability of Italians museums. To do that, the authors used MPADI2 index based for 13 annual reporting about 93 Italian museums. The results were compared with the UK, USA and European museums annual reporting and showed that have different disclosing practices across countries. Ong et al. (2016) examined the relationship between environmental disclosures and financial performance of public listed companies in Malaysia. The authors concluded that only the quality of environmental disclosure has the positive relationship with the company's earnings per share.

2.2   Linguistic obfuscation

It is the use of language resources to obfuscate or divert the attention. The study of narratives in linguistics was initially developed by Beaugrande and Dressler (1981). In the theory of narrative communication, they defined a set of seven principles determining the effectiveness of communication, namely: cohesion, coherence, acceptability, intentionality, informativeness, situationality, and intertextuality. Reeves et al. (2005) argue that the different narrative documents published by the companies are related to the different intentions thereof and that such intentions are correlated with the linguistic choices of agents. According to Rutherford (2003), the financial accounts are presented as a challenge even greater for regulators than for the accounting figures themselves.

Some studies have been carried out to verify the use of linguistic resources with the aim to obfuscate information on narrative accounting reports. Kong (2001) investigated on how to identify the obfuscate and linguistic manipulation in MR. Results showed that the organisational marketing section is constructed in a sophisticated way and there is an appeal to the intertextuality. Rutherford (2003) examined whether the companies conducted misuse of textual complexity in order to obfuscate the financial and operating information, thus preventing the effective communication and better governance practices. The author found no significant relationship between the use of textual complexity and the companies' features or their poor performance. Moffitt and Burns (2009) evidenced that companies classified as fraudulent used in their narrative more complex words and resorted for those with a sense of 'action' and 'cause'. Hassan (2014) examined how stakeholders create societal expectations related to risk disclosures in financial institutions. They found that financial institutions utilise 'assertive' and 'defensive' disclosure tactics to gain, maintain and restore their social legitimacy. Evans and Pierpoint (2015) have analysed annual report narratives and found that the narratives are active in discharge the director's accountability by demonstrating their work as socially and economically useful and managing expectations. Wickham et al. (2015) tested marketing priorities in the annual reports. The results show that the communication of marketing strategies has a central role in the annual reporting and that there are several organisation's marketing strategies that must be communicated to internal and external stakeholder groups. Moreno and Casasola (2015) focused on the readability evolution of annual report narratives applying an adapted version of the Flesch readability formula. They concluded that a range of variables that might influence readability, including profitability. Iatridis (2016) verified that pessimistic language in annual reports is higher for companies with lower earnings manipulation and higher power. More, the authors

concluded that annual reports tend to be more pessimistic and tend to guide analysts in the way to reduce the magnitude of a negative market reaction to underperformance.

2.3 Similarity

For Hatzivassiloglou et al. (1999), two text units are similar when they share the same focus on a concept, actor, object, or common action. For Choi et al. (2001), the metric similarity estimates the probability of two segments to describe the same matter. According to the authors, there are two methods for estimating the similarity: the lexical cohesion method developed by Halliday (1976), which detects parts of the text with similar terms; and the word repetition method, which detects the sequence of words presented. Uzuner et al. (2004b) emphasise that to assess the similarity between two documents it is necessary to analyse both the content and the expression presented in the documents. The authors define content as the facts reported in the document, while the term expression is linked to the authors' language preferences when presenting such facts (contents). For Frawley (1992 apud Uzuner et al., 2004b), there are two ways of assessing the similarity: by the equality and difference between the texts. Therefore, there are two types of similarity: representational (or syntactical) and behavioural (or semantical). The syntactical similarity provides the text as a sequence of characters forming a complex textual structure. Thus, this type of similarity can be determined as to the shape, the representation properties or characteristics; and its analysis is performed by examining blocks, sentences and text type. As to the other type, the semantical similarity, two texts are considered similar when the similarity is in its meaning; worth mentioning that it is harder to capture such similarities (Walenstein et al., 2007).

Uzuner et al. (2004a) state that there are several events on which it is useful to determine whether the contents of the two texts are similar, including the detection of plagiarism and copyright violations. Notwithstanding its relevance, the study of textual similarity has been forgotten by scientific community (Uzuner et al., 2004b). In this sense, some researches are being developed in the Science and Information Programming area aimed at developing methodologies for the similarity analysis between two documents. Hatzivassiloglou et al. (1999) presented a metric similarity methodology combining information from several linguistic indicators in order to measure the semantic distance between pairs of text units. Walenstein et al (2007) presented similarity measurement templates made by computer programs. Lin et al. (2013) proposed a new computing model of spam text similarity based on three kinds of Chinese text similarity calculation methods. Boubacar (2014) developed a way to measure similarity using words occurrence. For him, this methodology can be used for a wide range of content analysis tasks based on semantic distance.

3 Research design

The object hereof is Brazilian firms with stocks traded at stock exchanges. The Brazilian Law state that these firms shall disclosure documents representing the set of information towards the many accounting users and that the MR is aimed at evidencing the main facts administrative facts of the year ended. The Securities and Exchange Commission (CVM), governing body of the capital market in Brazil, addresses these matters in their standards.

It is worth noting that some of the items recommended by CVM induce the similarity, such as the description of businesses, products and services; while others motivate changes in the information disclosed throughout the years, such as the comments on the general economic situation. Brazil adopted the IASB international accounting standards and such matter is also the subject thereof, highlighting how important it is that the information contains quality standards.

Brazilian firms' MR for the period from 2008 and 2009 have been collected, one by one, at CVM location. The sample selection took into account the availability of MR and the availability of economic/financial data. The final study sample was 311 companies, corresponding to 57% of listed companies.

After collection, the MR underwent a textual similarity analysis. The methodology used for the analysis was to assess the similarity by detecting the differences between the two texts. For this end, a content analysis software has been used, called ExamDiff Pro, which may be used for analysing texts in Portuguese language. This tool allows to compare text files, highlighting differences up to the level of lines, words and characters. For the purposes hereof, the syntactical similarity was used. The program also performs a widespread match in order to be allowed to align the similar parts of the text and automatically detect the differences between the compared text files. After comparing text files, ExamDiff Pro shows three different statistics: total differences between the texts, line (or sentences) differences and intra-line differences (or words). The 'total difference' is a weighting performed by the program of the 'line differences' and the 'Intra-line differences'. The line and intra-line differences statistics, in turn, are the result of the sum of the number of sentences or words, as appropriate, which have been added, deleted and modified. Therefore, we have the following equation: line differences = number of lines added + number of deleted lines + number of modified lines. Intra-line differences calculation is obtained in the same way.

Fields and lines left blank have not been considered, as well as illustrations. Furthermore, the investigation of textual similarity shall only be performed for documents in .doc format. The comparison was performed for four similarity levels: 100, 90, 60 and 50. Additionally, the number of characters, words, sentences, and the Flesch legibility value were obtained in each report present in the sample.

Prior to the establishment of the similarity indicators, it was necessary to carry out a *t*-test for the three similarity statistics provided by the program, namely: total differences, line differences and intra-line differences. The test sought whether the four degrees of similarity considered (100, 90, 60 and 50) influenced the explanatory content. Results showed that this was not the case, except for the intra-line differences, between level 60 and 100 (p-value = 0.02).

Based on the results of *t*-test were created indicators by crossing the variables obtained from the software (for the levels 100 and 60) with those provided by the text program. This procedure proved to be necessary due to the scale issue. Figure 1 provides an example on how the indicators have been created.

Given that the performed resulted in reasonable amount of variables (nine per similarity level), the calculation of correlation matrix was performed, aimed at verifying the interactions between indicators, thus reducing them to a sole Similarity indicator considering the level 100 and the level 60. Pearson (parametric) and Spearman (non-parametric) correlation coefficients were calculated considering 95% of confident interval. In both cases the correlations were high and positive. With that in mind, only one indicator of each level was chosen to be used for developing the regression model, in

this case, the 'intra-line differences'. However, as in the average test, it was found that for the 'intra-line differences' statistics have a higher difference between the two levels (100 and 60) provided.

Figure 1 Exemplification of the development of indicators
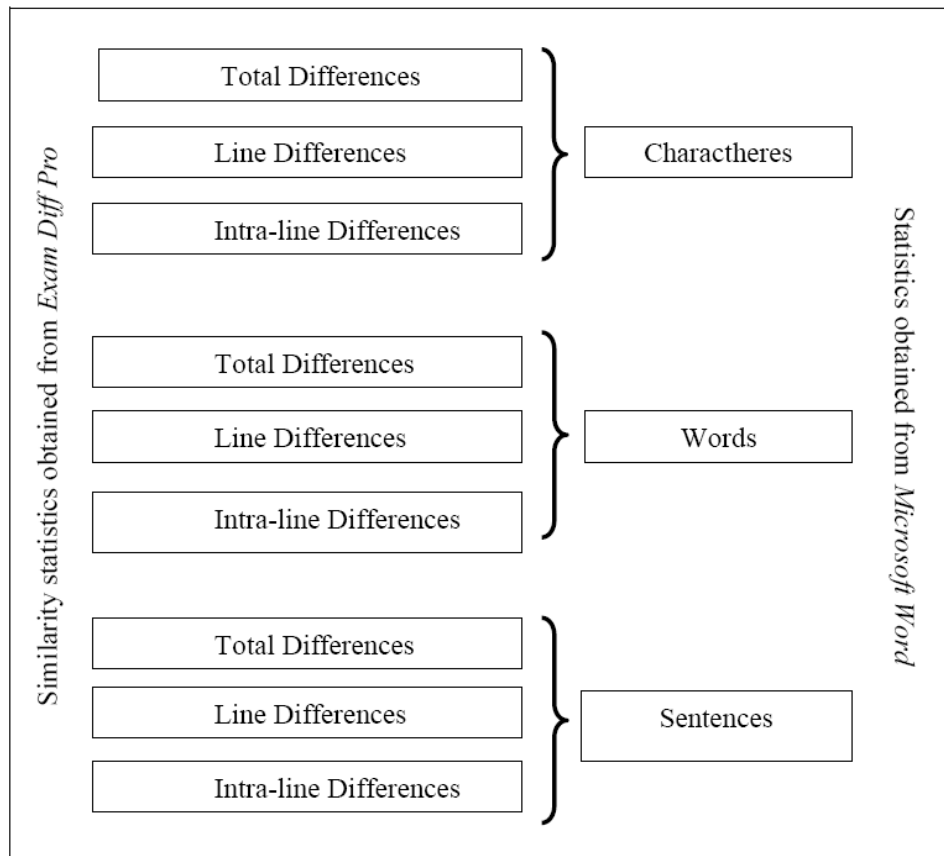
Subsequently, in order to determine the influence of characteristics:

1   language complexity

2   organisational complexity

3   financial performance

4   risk

5   political cost

6   size of the company in the similarity indicator of the MR and to test the other hypothesis listed in this research, a multiple linear regression technique was used.

In this sense, the independent variables contained in Table 1 were used.

Table 1   Independent variable

Variable	Measurement proxy	Name	Description	Type
Organisational complexity	Business sector	SECTOR	Set of dummy variables that determine the business sector of the company: public utility, financial and others, industrial goods, basic materials, construction and transport, basic materials, cyclical consumption, non-cyclical consumption, information technology, telecommunications, oil and gas, unclassified	Dummy
	Level of corporate governance	GC	Set of dummy variables that determine whether the company is involved in some levels of BMF&BOVESPA governance; 0 = no and 1 = yes	Dummy
	Press	IP	Number of press reports of the specialised press in the period from January 1, 2008 to December 31, 2009, according to Factiva database	Metric
	Trade in foreign exchange	ADR	Set of dummy variables that determine whether the company has securities traded on US exchanges (American Depositary Receipts)	Dummy
	Ownership concentration	CONCENT	Percentage share of the largest shareholder in the capital of the company	Metric
	Auditing	AUDITING	Set of dummy variables that determine whether or not the company has been audited by a Big Four in 2008; 0 = no and 1 = yes	Dummy
Financial performance	Net book value variable	DΔPL	Set of dummy variables that determine whether the equity in 2009 have increased or decreased compared to 2008; 0 = decrease and 1 = increase	Dummy
	Profitability variation	DΔLL	Set of dummy variables that determine whether the net income in 2009 have increased or decreased compared to 2008; 0 = decrease and 1 = increase	Dummy
	Revenue variation	DΔREC	Set of dummy variables that determine whether the Net operating revenue in 2009 have increased or decreased compared to 2008; 0 = decrease and 1 = increase	Dummy
	Market liquidity	LIQMERC	Ratio of the trading volume in 2008 and the net equity in 2008	Metric
	Financial independence	FI	Ratio of the net equity and the total asset on 31/12/2008	Metric
	Current liquidity	CL	Ratio of the current assets and liabilities on 31/12/2008	Metric
Language complexity	Variation in the size of MR	ΔMR	Difference in the size of the 2009 MR (in characters) and the size of the 2008 MR	Metric
	Variation in the size of MR	DΔMR	Set of dummy variables that determine whether the size (in characters) of the 2009 MR has increased or increased compared to 2008; 0 = decrease and 1 = increase	Dummy
	Readability	READ	Easy reading	Metric
Risk	Beta	BETA	Statistical measure that reflects the risk of an asset as to the risk of the market portfolio	Metric
Political costs	Regulated sector	REGS	Set of dummy variables that determine whether or not the company belongs to a sector strictly regulated by the government; 0 = no and 1 = yes. Regulated sectors: public utility, financial and others; oil and gas and telecommunications	Dummy
Company size	Size	SIZE	Asset value natural logarithm (in thousand reais) on 31/12/2008	Metric

The analysis of the influence of the independent variables in the similarity indicator was performed by total and by group. At first, the multiple regression technique was applied to all companies selected in the sample for the level 100 and 60. Then, the indicator of each level was divided in ascending order of similarity and segregated into high, medium and low tertiles similarity. Later, the aggregate indicator of high value tertiles and low similarity indicators were compared with the respective linguistic and organisational complexity features, the financial performance, the risk, the political cost and the size of companies within each tertile.

For the selection of variables belonging to the models, it was adopted the stepwise method at a confident interval level of 95%. In addition, the following robustness tests were carried out for the calculated models. As a result of the multiple regressions, it is expected to find a positive ratio ($\beta > 0$) between the similarity index and the variables related to the language complexity, the financial performance, the risks, the political costs, and the company size. In the event such a result is observed, the administrators shall demonstrate their concern in evidencing the administration achievements, as will be evidenced an attempt by managers to justify to the shareholders and other stakeholders the increases in activity levels. For the variables related to the organisational complexity, we believe that the involvement of the company in the levels of corporate governance, the securities trading in the foreign markets and the fact that companies are audited by one of the four largest auditing firms worlds triggers an increase in the information available to investors of such companies.

However, there is an understanding that the higher the ownership concentration and the investment risk offered by the company, the smaller the differences between MR disclosed from one period to another ($\beta < 0$). It is also understood that the higher the risk attributed to the company, the less similar their reports should be, since such a fact would reduce the information asymmetry and, therefore, the uncertainty for investors.

4 Main results

Table 2 shows the descriptive statistics of the number of lines (sentences) and intra-lines (words) added, deleted and modified compared to the 2008 and 2009 MR for the similarity levels of 60 and 100, respectively.

It is observed that the average number of modified lines is significantly higher than the average number of added lines (approximately 52 times higher). It is important to note that the change statistics is calculated when the sentence structure is maintained and the changes are limited to some words. The reports of the companies Sergen S.A and Abyara Planejamentos Imobiliários S.A are an example of these findings. In the first, no text line has been added or deleted and the modifications thereof are limited to changes within the line, i.e., in the words. In the second case there were significant changes and 74 additional lines were detected.

In addition, when analysing the table of added and deleted sentences frequency for the levels 60 and 100, it was found that approximately 70% of the companies under study had not add or deleted any sentence. Therefore, the results hereof support the hypothesis that most companies tend to maintain a standard MR.

Table 2   Descriptive statistics

		N	Minimum	Maximum	Average	Standard deviation	Asymmetry	Kurtosis
Level 60	LADIC	311	0	465	7.60	2.09	8.79	92.72
	ILADIC	311	0	1,508	197.21	12.28	2.28	7.16
	LDEL	311	0	81	3.43	0.61	4.44	22.27
	ILDEL	311	0	1,644	182.95	12.18	2.77	10.98
	LMUD	311	0	3,876	364.05	23.20	3.13	18.47
	ILMUD	311	0	7,868	905.98	58.22	2.69	10.42
Level 100	LADIC	311	0	465	7.65	2.09	8.79	92.69
	ILADIC	311	0	1,521	207.22	12.82	2.28	6.88
	LDEL	311	0	81	3.43	0.61	4.44	22.27
	ILDEL	311	0	1,609	193.50	12.67	2.61	9.29
	LMUD	311	0	3,876	366.13	23.16	3.14	18.57
	ILMUD	311	0	7,650	922.58	59.25	2.60	9.36

Notes: Where: LADIC = added lines; LDEL = deleted lines; LMUD = changes lines
ILADIC = intra-line added; ILDEL = intra-line detected; ILMUD = intra-line
changed.

4.1   Relation between independent variables and the similarity indicator, total

4.1.1   Similarity level 100

As a result of this regression model, only the following variables returned significant results: variation of MR size; size; readability, shareholders' equity variation, corporate governance, telecommunications and basic materials sectors. With regards to the test of model existence, it was obtained a coefficient F in the amount of 10.33, significant at 5%. In this regard, the hypothesis of regression is accepted. Results are shown in Table 3.

For the readability variable was found a β > 0, as expected, evidencing that the easy reading is positively related to the increase of differences in the narrative reports (or decrease in similarity). The PL variation indicates that when the equity increases, there is an increase in similarity. The corporate governance variable proved to be relevant, indicating that when a company has best corporate governance practices the level of similarity of MR decreases. The size also showed β > 0; thus, the larger the company size, the smaller the similarity, supporting the hypothesis that larger companies – given their greater visibility in the market and society – have the concern to provide more information to justify their activities and obfuscate their performance information.

It was also verified that the basic materials sector, which includes pulp and paper, pesticides and fertilisers companies, tends to show less similar narrative reports. The result may be justified, once companies in these sectors should have a greater concern to quantify and provide information related to environmental issues. Similarly, the telecommunications sector was also relevant and with β > 0. This sector is considered as regulated and subject to constant inspections by the Brazilian Government. In general, the fact that other independent variables were not explanatory for this similarity level may suggest that the maintenance of similar reporting structures is a common practice in Brazilian companies studied in the sample.

Table 3 Variables that explain the 100 similarity indicator, per total

Variable	Predictor variables	β	Stand. β	t-test	Sig.	VIF	Adjusted R^2	ε	Test F	Sig.
	Constant	6.20		2.57	0.01		0.18	5.41	10.33	0.03
Language complexity	DΔMR	3.24	0.27	5.17	0.00	1.03	0.18	5.41	10.33	0.03
	READ	0.32	0.12	2.21	0.03	1.17				
Financial performance	DΔPL	–0.14	–0.20	–3.76	0.00	1.04	0.18	5.41	10.33	0.03
Organisational complexity	GC	3.19	0,18	3.45	0.00	1.04	0.18	5.41	10.33	0.03
	STELEC	1,70	0.14	2.53	0.01	1.18				
	SMATBAS	4.20	0.13	2.32	0.02	1.08				
Company size	SIZE	1.60	0.12	2.17	0.03	1.08	0.18	5.41	10.33	0.03

Notes: Model: $IS_i = \beta_1 + \beta_2 Sector_i + \beta_3 GC_i + \beta_4 Size_i + \beta_5 CL_i + \beta_6 IP_i + \beta_7 ADR_i + \beta_8 Concent_i + \beta_9 Auditing_i + \beta_{10} Read_i + \beta_{11} D\Delta MR_i + \beta_{12} D\Delta PL_i + \beta_{13} D\Delta LL_i + \beta_{14} D\Delta Rec_i + \beta_{15} LiqMerc_i + \beta_{16} Beta_i + \beta_{17} FI_i + \varepsilon.$
Where: DΔMR = variation dummy of the MR size; READ = readability 2008; DΔPL = dummy variation of shareholders' equity; GC = corporate governance; STELEC = telecommunication sector; SMATBAS = basic material sector; SIZE = company size; IS = company similarity indicator i; Bn = n variable coefficient for the company i; ε = regression error instrument.

4.1.2 Similarity level 60

In the second regression model calculated, only five variables returned as significant, namely: variation of MR size, size, readability, shareholders' equity variation, and basic materials sector. With regards to the regression existence test, the coefficient F presented the amount of 9,813, significant at 5%; therefore, the hypothesis of the existence of the model is accepted, as shown in Table 4.

Table 4 Variables that explain the similarity indicator 60, per total

Variable	Predictor variables	β	Stand. β	t-test	Sig.	VIF	Adjusted R^2	ε	Test F	Sig.
	Constant	5.11		1.94	0.05		0.13	6.09	9.81	0.03
Language complexity	DΔMR	2.76	0.21	3.92	0.00	1.02	0.13	6.09	9.81	0.03
	READ	–0.11	–0.14	–2.64	0.01	1.01				
Financial Performance	DΔPL	2.15	0.14	2.60	0.01	1.06	0.13	6.09	9.81	0.03
Organisational complexity	SMATBAS	4.07	0.21	3.93	0.00	1.03	0.13	6.09	9.81	0.03
Company size	SIZE	0.34	0.12	2.21	0.03	1.04	0.13	6.09	9.81	0.03

Notes: Model: $IS_i = \beta_1 + \beta_2 Sector_i + \beta_3 GC_i + \beta_4 Size_i + \beta_5 CL_i + \beta_6 IP_i + \beta_7 ADR_i + \beta_8 Concent_i + \beta_9 Auditing_i + \beta_{10} Read_i + \beta_{11} D\Delta MR_i + \beta_{12} D\Delta PL_i + \beta_{13} D\Delta LL_i + \beta_{14} D\Delta Rec_i + \beta_{15} LiqMerc_i + \beta_{16} Beta_i + \beta_{17} FI_i + \varepsilon.$
Where: DΔMR = variation dummy of the MR; READ = readability 2008; DΔPL = dummy variation of PL; SMATBAS = basic material sector; SIZE = company size.

By analysing Table 4, it is observed that these same variables proved to be relevant in the previous model. However, readability variables and the shareholders' equity variation showed an opposite sign to the previously obtained. In this case, an increase in the readability of reports triggers an increase in MR similarity. Likewise, an increase in PL triggers a decrease in the MR similarity in subsequent periods.

4.2   Relation between independent variables and the similarity indicator, tertile

4.2.1   First tertile – similarity level 100

The results of the regression model for the group with the lowest similarity indicators returned as relevant the variable: net profit variation, beta, audit, MR size variation and construction and transport sectors, as shown in Table 5.

Table 5   Variables that explain the similarity indicator 100, per tertile (first tertile)

Variable	Predictor variables	β	Stand. β	t-test	Sig.	VIF	Adjusted R²	ε	Test F	Sig.
	Constant	4.89		8.91	0.00		0.65	0.85	9.48	0.00
Language complexity	DΔMR	1.01	0.29	2.14	0.05	1.29	0.65	0.85	9.48	0.00
Financial Performance	DΔLL	1.21	0.42	2.99	0.01	1.39	0.65	0.85	9.48	0.00
Operational complexity	AUDIT	−1.27	−0.37	−2.67	0.02	1.21	0.65	0.85	9.48	0.00
	SCONST	−2.49	-0.77	−5.28	0.00	1.24				
Risk	BETA	1.44	0.34	2.50	0.02	1.24	0.65	0.85	9.48	0.00

Notes: Model: $IS_i = \beta_1 + \beta_2 Sector_i + \beta_3 GC_i + \beta_4 Size_i + \beta_5 CL_i + \beta_6 IP_i + \beta_7 ADR_i + \beta_8 Concent_i + \beta_9 Auditing_i + \beta_{10} Read_i + \beta_{11} D\Delta MR_i + \beta_{12} D\Delta PL_i + \beta_{13} D\Delta LL_i + \beta_{14} D\Delta Rec_i + \beta_{15} LiqMerc_i + \beta_{16} Beta_i + \beta_{17} FI_i + \varepsilon$.
Where: DΔMR = variation dummy of the MR; DΔLL = net profit variation dummy; SCONST = construction and transport sectors; AUDIT = auditing; BETA = beta; IS = company I similarity indicator, in the group (tertile) j; Bn = n variable coefficient, for company i, in the group (tertile) j; ε = regression error instrument.

For this test, the performance variable net income variation proved to be significant and with β > 0, as expected, indicating that the profit increase results in an increase in the amount of information. The audit variable, with β < 0, shows that companies with the *most similar* MR, mostly, have not been audited by one of the big four. The risk showed a positive coefficient; the conclusion is that companies with more similar MR appear as higher investment risks. Since the price of any asset is calculated based on the several information publicly available to investors, in the cases where the information is reduced the asymmetry of information increases the risk and uncertainty of those investing funds in the company. The construction and transport sector had a negative slope; the majority of companies with the most similar MR do not comprise these sectors. It is deducted that there is a concern by the managers of companies in these sectors to disclose recent achievements, as well as to justify high results achieved due to the high growth rates effective in recent years, with the recovering of construction and transportation sectors.

4.2.2 Third tertile – similarity level 100

The results of the regression model for the group with the highest similarity indicators returned as relevant only the press variable. The results are shown in Table 6.

Table 6 Variables that explain the Similarity Indicator 100, per tertile (third tertile)

Variable	Predictor variables	β	Stand. β	t-test	Sig.	VIF	Adjusted R^2	ε	F test	Sig.
	Constant	11.73		25.54	0.00		0.46	2.37	26.28	0.00
Organisational complexity	IP	0.00	0.69	5.13	0.00	1.00	0.46	2.37	26.28	0.00

Notes: Model: $IS_i = \beta_1 + \beta_2 Sector_i + \beta_3 GC_i + \beta_4 Size_i + \beta_5 CL_i + \beta_6 IP_i + \beta_7 ADR_i + \beta_8 Concent_i + \beta_9 Auditing_i + \beta_{10} Read_i + \beta_{11} D\Delta MR_i + \beta_{12} D\Delta PL_i + \beta_{13} D\Delta LL_i + \beta_{14} D\Delta Rec_i + \beta_{15} LiqMerc_i + \beta_{16} Beta_i + \beta_{17} FI_i + \varepsilon$.
Where: IP = number of new published by specialised press

The higher the number of news published in the specialised press, the greater the amount of incremental information. This may be justified by the greater need of the administration to confirm and/or to attempt to obfuscate information released by the press.

4.2.3 First tertile – similarity level 60

For the regression model for indicators with lower similarity indicators (first tertile) referring to level 60, no variable proved to be relevant.

4.2.4 Third tertile – similarity level 60

As to the companies with greater similarity indicators (third tertile), the only relevant variable was the shareholders' equity variation, as show in Table 7.

Table 7 Variables that explain the similarity indicator 60, per tertile (third tertile)

Variable	Predictor variables	β	Stand. β	t-test	Sig.	VIF	Adjusted R^2	ε	F test	Sig.
	Constant	15.93		12.21	0.00		0.19	2.92	7.39	0.01
Financial performance	D∆PL	–3.91	–0.47	–2.72	0.01	1.00	0.19	2.92	7.39	0.01

Notes: Model: $IS_i = \beta_1 + \beta_2 Sector_i + \beta_3 GC_i + \beta_4 Size_i + \beta_5 CL_i + \beta_6 IP_i + \beta_7 ADR_i + \beta_8 Concent_i + \beta_9 Auditing_i + \beta_{10} Read_i + \beta_{11} D\Delta MR_i + \beta_{12} D\Delta PL_i + \beta_{13} D\Delta LL_i + \beta_{14} D\Delta Rec_i + \beta_{15} LiqMerc_i + \beta_{16} Beta_i + \beta_{17} FI_i + \varepsilon$.
Where: D∆PL = shareholders' equity variation.

For those companies with less similar reports, it relates to the decrease in the shareholders' equity value. It can be concluded that there is an attempt by managers to justify a possible performance decrease.

4.3   Summary of research hypotheses and results found

Based on the statistics of the sentences and words added, deleted and modified, it was found that over 70% of Brazilian companies reproduced the MR structure in the subsequent years. In addition, the reports of these companies maintained the same amount of information, with the only changes being the dates and values (Hypotheses 1 and 2) (Table 8).

Table 8   Summary of research hypotheses and results found

Research hypotheses	Description	Accepted	Rejected	Partially confirmed
1	Companies tend to reproduce the structure of their MR over time.	X		
2	There are no significant changes in the amount of information disclosed in the MR during the study period.	X		
3	Depending on their size, companies tend to present MR roughly similar in subsequent periods.	X		
4	Companies within regulated sectors (higher political costs) tend to have less similar MR.			X
5	Companies with better performance indicators tend to have less similar MR.		X	
6	Companies with more complex organisational structures tend to have less similar MR.			X
7	Companies with higher ownership concentration indicators tend to have more similar MR.		X	
8	Companies representing higher investment risks tend to disclose more similar MR.		X	

For the regression models per total analysed, the similarity indicator was influenced by the size of the company. In this sense, the size of the company is directly related to the disclosure of less similar MR (Hypothesis 3). With regards to Hypothesis 4, which assumes that companies within regulated sectors tend to have less similar accounting narratives, this hypothesis was partially confirmed: only one of the models presented as significant one of the regulated sectors. In addition, the regulated sector variable, which covers all sectors regulated in the model, was not presented as relevant in any of the tests performed.

Hypothesis 5 argues that companies with better financial performance indicators presented less similar MR. However, this hypothesis was rejected. First because only one of the performance indicators proved to be relevant (the shareholders' equity variation), but with different slope signals.

With regards to Hypothesis 6 that companies with more complex organisational structures would present MR with lower similarity indexes, it was partially confirmed since the variable corporate governance proved to be relevant, and only in one of the models analysed.

Hypothesis 7, which provides that companies with higher ownership concentration tend to have more similar MR, cannot be confirmed, the same applies for Hypothesis 8.

5   Final considerations

This study proposes an investigation in order to verify whether there is duplication in information management discloses from a financial year to another, thus identifying which variables explain the similarity.

The results found showed that over 70% of Brazilian companies in the sample did not show relevant changes in their MR. Most companies made only minor changes such as date, names and values. This result confirms the following hypotheses hereunder: (1) the Brazilian companies reproduce certain structures in their MR in subsequent years; and (2) there were no significant changes regarding the amount of information contained in the reports. By confirming this hypothesis, this study demonstrates that Brazilian companies may be failing to contribute to users, and misleading them to possible errors. Moreover, the companies also may be missing a great opportunity to spread its achievements to the market.

In addition, it was found that the size of the company has a positive impact on the change of the information contained in the reports (3). Therefore, the larger the company, the lower the similarity of its MR in subsequent years. It can be observed that there is a concern by the large companies' administrators to present their management, by changing and/or including information in their narratives. Furthermore, the results found are consistent with the literature that larger companies, for having high political costs and high visibility in the market, provide reports in order to justify the higher level of their activities.

Is worth mentioning that three hypotheses raised could not be proven in this study and, therefore, were rejected, namely: (5) companies with better performance indicators tend to present less similar reports; (7) companies with higher levels of ownership concentration tend to have more similar reports; and (8) companies representing a higher investment risk have more similar MR. This study failed to prove, through the tests performed, whether the companies within the regulated sectors (4) and those with more complex organisational structures (6) tend to present less similar reports for the analysed period.

The lack of confirmation of part of the hypotheses raised hereof may be due to the study period. The short period analysed is a limiting factor of the research. However, this limitation was necessary, since the analysis of additional periods would result in an exponential increase of the study indicators. Another limiting factor of the research is related to the inability to control the macroeconomic variables affecting the information in the accounting narratives. As well as to the limitations linked to the type of similarity measurement used. The syntactical similarity lists two texts analysing a sequence of words, therefore, it measures only the differences between the words presented and their quantities, thus not contemplating whether these differences have the same semantic value.

It is important emphasising that the method used may be applied to assess the similarity in other instruments aimed at the disclosure to the market, such as the public offer prospectus, appraisal reports of companies, relevant facts and even reference forms.

References

Abrahamson, E. and Amir, E. (1996) 'The information content of the president's letter to shareholders', *Journal of Business Finance & Accounting*, Vol. 23, No. 8, pp.1157–1182.

Bambagiotti-Alberti, L., Manetti, G. and Sibilio-Parri, B. (2014) 'The quality of annual reporting by Italian museums: an international comparison', in *From Theory to Practice: The Bridge to the Next Generation of Performance Management and Measurement*, Pavia, 6–7 November 2014, Sidrea, pp.1–24.

Beattie, V., McInnes, B. and Fearnley, S. (2004) 'A methodology for analysing and evaluating narratives in annual reports: a comprehensive descriptive profile and metrics for disclosure quality attributes', in *Accounting Forum*, September, Vol. 28, No. 3, pp.205–236.

Beaugrande, R. and Dressler, W.U. (1981) *Einführung in die Textlinguistik*, Vol. 28, Niemeyer, Tübingen.

Boo, E.F. and Simnett, R. (2002) 'The information content of management's prospective comments in financially distressed companies: a note', *Abacus*, Vol. 38, No. 2, pp.280–295.

Boubacar, A. (2014) 'Valuing semantic relatedness', in *Information Technology and Artificial Intelligence Conference (ITAIC), 2014 IEEE 7th Joint International*, IEEE, December, pp.1–5.

Bryan, S.H. (1997) 'Incremental information content of required disclosures contained in management discussion and analysis', *The Accounting Review*, April, Vol. 72, No. 2, pp.285–301.

Choi, F.Y.Y., Hasting, P.W. and Moore, J. (2001) 'Latent semantic analysis for text segmentation', *Computational Linguistics*, Vol. 32, No. 1, pp.5–12.

Courtis, J.K. (1998) 'Annual report readability variability: tests of the obfuscation hypothesis', *Accounting, Auditing & Accountability Journal*, Vol. 11, No. 4, pp.459–472.

Dhanani, A. and Connolly, C. (2015) 'Non-governmental organizational accountability: talking the talk and walking the walk?', *Journal of Business Ethics*, Vol. 129, No. 3, pp.613–637.

Dye, R.A. (2001) 'An evaluation of 'essays on disclosure' and the disclosure literature in accounting', *Journal of Accounting and Economics*, Vol. 32, No. 1, pp.181–235.

Evans, L. and Pierpoint, J. (2015) 'Framing the Magdalen: sentimental narratives and impression management in charity annual reporting', *Accounting and Business Research*, Vol. 45, Nos. 6–7, pp.661–690.

Halliday, M.A. (1976) *System and Function in Language: Selected Papers*, Oxford Press, London.

Hassan, M.K. (2014) 'Risk narrative disclosure strategies to enhance organizational legitimacy: evidence from UAE financial institutions', *International Journal of Disclosure and Governance*, Vol. 11, No. 1, pp.1–17.

Hatzivassiloglou, V., Klavans, J.L. and Eskin, E. (1999) 'Detecting text similarity over short passages: exploring linguistic feature combinations via machine learning', in *Proceedings of the 1999 Joint Sigdat Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, June, pp.203–212.

Herremans, I.M. and Ryans, J.K. (1995) 'The case for better measurement and reporting of marketing performance', *Business Horizons*, Vol. 38, No. 5, pp.51–60.

Iatridis, G.E. (2016) 'Financial reporting language in financial statements: does pessimism restrict the potential for managerial opportunism?', *International Review of Financial Analysis*, Vol. 45, pp.1–17.

Jones, M.J. and Shoemaker, P.A. (1994) 'Accounting narratives: a review of empirical studies of content and readability', *Journal of Accounting Literature*, Vol. 13, No. 1, p.142.

Kong, K.C. (2001) 'Marketing of belief: intertextual construction of network marketers' identities', *Discourse & Society*, Vol. 12, No. 4, pp.473–503.

Lee, J. and Hong, Y.S. (2014) 'Business model mining: analyzing a firm's business model with text mining of annual report', *Industrial Engineering & Management Systems*, Vol. 13, No. 4, pp.432–441.

Lin, H., Wang, J. and Xu, C. (2013) 'A joint computing model based on Chinese spam text similarity', *International Conference on Services Science and Services Information Technology*.

Llewellyn, S. (1999) 'Narratives in accounting and management research', *Accounting, Auditing & Accountability Journal*, Vol. 12, No. 2, pp.220–237.

Marston, C.L. and Shrives, P.J. (1991) 'The use of disclosure indices in accounting research: a review article', *The British Accounting Review*, Vol. 23, No. 3, pp.195–210.

Moffitt, K. and Burns, M.B. (2009) 'What does that mean? Investigating obfuscation and readability cues as indicators of deception in fraudulent financial reports', *AMCIS 2009 Proceedings*, p.399.

Moreno, A. and Casasola, A. (2015) 'A readability evolution of narratives in annual reports a longitudinal study of two Spanish companies', *Journal of Business and Technical Communication*, Vol. 30, No. 2, pp.202–235.

Ong, T.S., Tho, H.S., Goh, H.H., Thai, S.B. and Teh, B.H. (2016) 'The relationship between environmental disclosures and financial performance of public listed companies in Malaysia', *International Business Management*, Vol. 10, No. 4, pp.461–467.

Reeves, T.C., Ford, E.W., Duncan, W.J. and Ginter, P.M. (2005) 'Communication clarity in strategic management data sources', *Strategic Organization*, Vol. 3, No. 3, pp.243–278.

Rutherford, B.A. (2003) 'Obfuscation, textual complexity and the role of regulated narrative accounting disclosure in corporate governance', *Journal of Management and Governance*, Vol. 7, No. 2, pp.187–210.

Springer, L. (1992) 'Enhancing the annual report: investor relations and the MD&A', *The Journal of Bank Accounting & Auditing*, Vol. 5, No. 2, pp.27–29.

Sydserff, R. and Weetman, P. (1999) 'A texture index for evaluating accounting narratives: an alternative to readability formulas', *Accounting, Auditing & Accountability Journal*, Vol. 12, No. 4, pp.459–488.

Uzuner, O., Davis, R. and Katz, B. (2004a) *Recognizing Text Similarity*, Massachusetts Institute of Technology.

Uzuner, O., Davis, R. and Katz, B. (2004b) 'Using empirical methods for evaluating expression and content similarity', in *Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 2004*, January, IEEE, p.8.

Van Geyt, D., Van Cauwenberge, P. and Vander Bauwhede, H. (2014) 'Does high-quality corporate communication reduce insider trading profitability?', *International Review of Law and Economics*, March, Vol. 37, pp.1–14.

Walenstein, A., El-Ramly, M., Cordy, J.R., Evans, W.S., Mahdavi, K., Pizka, M. and von Gudenberg, J.W. (2007) 'Similarity in programs', in *Dagstuhl Seminar Proceedings*, Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Watts, R.L. and Zimmerman, J.L. (1986) *Positive Accounting Theory*, Prentice Hall, New Jersey.

Wickham, M., Lehman, K. and French, L. (2015) 'Communicating marketing priorities in the not-for-profit sector: a content analysis of Australian state-museums' annual reports', *Corporate Ownership & Control*, Vol. 12, No. 2, pp.318–328.

Yuthas, K., Rogers, R. and Dillard, J.F. (2002) 'Communicative action and corporate annual reports', *Journal of Business Ethics*, Vol. 41, Nos. 1–2, pp.141–157.

Zeller, T.L., Stanko, B.B. and Jin, H. (2012) 'Investigating presentational change in company annual reports: an extension', *Academy of Accounting and Financial Studies Journal*, Vol. 16, No. 3, p.1.

Notes

1   Costs arising out of contracts and *lobbies*, as a result of the state regulation for the company activities (Watts and Zimmermam, 1986).