

Flows of information and meaning: a vocabulary approach to integrated thinking and reporting

Sonia Quarchioni

Department of Economics, Business and Statistics, University of Palermo, Palermo, Italy

Pasquale Ruggiero

Department of Business and Law, University of Siena, Siena, Italy and Brighton Business School, University of Brighton, Brighton, UK, and

Rodolfo Damiano

Department of Economics, Business and Statistics, University of Palermo, Palermo, Italy

Abstract

Purpose – Integrated reporting (IR) is increasingly becoming a practice useful not only for accountability but also for managerial purposes because of its potential role as a signifying practice for integrated thinking (IT). In this perspective, this paper aims to explore which of the objects that are represented in integrated reports provide materiality and common understanding to the concept of IT for its effective implementation within organizations.

Design/methodology/approach – This paper is based on a vocabulary approach for interpreting the texts of integrated reports as systems of words that are able to provide meaning for a common understanding of the concept of IT. In particular, by focusing on words and their relationships, the authors combine textual analysis and network text analysis to examine the structure of meaning embedded in the texts of integrated reports of five organizations, which serve as empirical cases for analysis during the period 2012-2018.

Findings – The concept of IT is dynamic in its meaning since in integrated reports it is represented by referring to different objects, in the case different types of capital (i.e. financial, human, social-relational, process, organizational and commercial), which are related to each other while following different paths over time. The dynamic nature of the meaning of IT affects the semantic orientation of the reports in a mutual relationship between IT (which conveys flows of information within the reports) and integrated reports (through which flows of meaning are made visible).

Originality/value – This paper opens the way to a linguistic approach for analyzing the different concepts related to IT to make them meaningful in creating (at least temporarily) a common



understanding, as well as facilitating coordination within organizations and between organizations and their environment.

Keywords Vocabularies, Textual analysis, Integrated reporting, Integrated thinking, Network text analysis

Paper type Research paper

1. Introduction

The increasing complexity of the business environment has affected organizations in three dimensions as follows: internal, contextual and transactional (Vasconcelos and Ramirez, 2011). Both managers and a growing number of stakeholders are interested in understanding how organizations face this complexity. In light of this, reporting practices could play an important role for organizations. Because of organizations' continuous search for legitimation in their environment through the adoption of and/or representation of socially acceptable value creation processes (DiMaggio and Powell, 1983; Modell, 2009; Monfardini *et al.*, 2013; O'Donovan, 2002), reporting practices shall increasingly abandon their compliance perspective to become managerial tools for understanding how such processes work and fit with the needs and requests of stakeholders (Barnabè *et al.*, 2019; Burritt and Schaltegger, 2010).

The release of the integrated reporting (IR) framework in 2013, by the International Integrated Reporting Council (IIRC), gave rise to a divisive debate on its possible role in improving corporate accountability (Brown and Dillard, 2014). However, especially during the last few years, the IIRC has stressed the idea that IR is no longer only a tool for accountability purposes viewing:

[...] reporting and thinking as two sides of the same coin – both necessary to enhance connectivity in the organization and enhanced communication on value creation. In combination, the benefits are rich and compelling: [...] Boards can actively use a multiple capitals approach to make strategic decisions about resource allocation and value creation over time, communicating the trade-offs between the capitals and the outcomes for business and society. This approach gives greater meaning to decisions, building trust inside the business and with stakeholders (International Integrated Reporting Council, 2016, p. 3).

According to the IR framework, integrated reports enhance accountability and performance management by representing organizations' use and transformation of different kinds of capital, as well as their interdependencies (International Integrated Reporting Council, 2013). Thinking about these various capitals and their interdependencies is a form of integrated thinking (IT) (Chartered Institute of Management Accountants, 2017; Deloitte, 2015). In this vein, IT is at the foundation of the IR process, which results periodically in the integrated report. Through this report organizations make IT material internally but also externally, being accountable within a cyclical activity through which organizations interact (directly and/or indirectly) with their stakeholders to agree on and spread meaning (International Integrated Reporting Council, 2016).

Along these lines, recent research has highlighted the narrative nature of integrated reports (Beattie, 2014; Higgins *et al.*, 2014) and their potential for specific forms of accountability toward investors and stakeholders, generally (Lai *et al.*, 2018). Although drawing on different perspectives, other studies have also suggested that reports containing financial and non-financial information can be the "practice" through which broader concepts can materialize and become meaningful (Busco *et al.*, 2018). Therefore, IR is a practice through which the expression of the organizations' *own* vision about their IT and value creation processes becomes especially evident and meaningful through the development of a common understanding while searching to fulfill the "mythical"

requirements of the framework (Gibassier *et al.*, 2018). From this perspective, integrated reports could be conceived of as a signifying practice, which conveys and materializes an organization's journey toward IT, by providing a meaning to this concept (Hall, 1997).

In particular, the IR practice makes the objects to which the concept of IT refers visible and material. However, there are still few insights into whether and how the meanings of IT on the part of organizations are projected through their integrated reports. Moreover, insights are needed to understand whether changes in these meanings can emerge from integrated reports over time. These issues deserve further investigation to grasp how organizations can use integrated reports to provide a common understanding of the concept of IT in its effective implementation, moreover, in light of the increasing complexity, which affects the relationship between organizations and their environment.

To fill this gap, the aim of this paper is to explore the objects through which the meaning of IT can be materialized and represented within integrated reports. Specifically, by adopting a longitudinal perspective this paper analyzes the integrated reports issued by first adopters who were awarded for the quality of their reports. This analysis aims to identify the capitals that are considered in the reports, and whether and how these capitals have been linked to each other to provide a representation of IT. Both the capitals and their linkages are the objects represented in the integrated reports providing meaning to the concept of IT.

To this purpose, this study is grounded on a vocabulary approach (Jones and Livne-Tarandach, 2008; Loewenstein *et al.*, 2012), according to which vocabularies are "conventions that people use as a common ground of knowledge and experiences to allow for meaningful communication and effective coordination" (Loewenstein *et al.*, 2012, p. 59), to interpret the texts of integrated reports as systems of words and their meaning regarding IT. In particular, by focusing on words and their relationships, we combine textual analysis (Lebart *et al.*, 1998) and network text analysis (Carley, 1994; Diesner and Carley, 2005) to examine the structure of meaning embedded in texts of integrated reports of five organizations, which serve as empirical cases for analysis, during the period 2012-2018.

From our findings, we argue that integrated reports are a signifying practice for making the concept of IT material and meaningful. Additionally, we emphasize the dynamic nature of the concept of IT because of its evolution over time due to the different prevalence of the capitals represented and due to change in the linking among them. The loss of weight in integrated reports of the financial and organizational capitals has been balanced by the growing weight of two other capitals, that is, commercial and socio-relational capitals. In addition, the connectivity of the multiple capitals represented increases over time. The dynamic nature of the meaning of IT affects the semantic orientation of the reports in a mutual relationship between IT (which conveys flows of information within the reports) and integrated reports (through which flows of meaning are made visible). In particular, the findings reflect an evolution of the semantic orientation of integrated reports, which are found to be more focused on relational aspects between organizations and their environments.

Our paper offers various contributions. First, we contribute to enriching the limited research on the concept of IT and its relationship to integrated reports (Feng *et al.*, 2017). While previous studies investigated the content of various sets of integrated reports to evaluate their level of relevance and quality by analyzing the focus of integrated reports on backward or forward looking information (de Villiers *et al.*, 2017), their extent of materiality by highlighting the most important content of the integrated reports for their preparers (Lai *et al.*, 2017) or their level of conciseness, integration and completeness in disclosing information (Busco *et al.*, 2019; Melloni *et al.*, 2017), they did not specifically investigate the meanings of the IT emerging from the reports. In this paper, we provide meaning to the concept of IT by singling out its objects of reference and the evolutionary path of

interactions between these objects. In so doing, we provide a dynamic perspective to the study of IT in a signifying practice, such as in IR.

Furthermore, we contribute more broadly to textual analysis research on corporate reporting (Brennan *et al.*, 2009; Milne *et al.*, 2009) by showing the importance of vocabulary structures to help explain latent meanings within texts. Therefore, we also make a methodological contribution by combining textual analysis with a network approach to discover the objects (in our case the capitals and their linkages) represented in integrated reports and consequently to materialize the concept of IT. In turn, this methodological contribution produces practical implications for the subjects involved in and/or interested in integrated reports, such as preparers, users and regulators. Preparers can have a better awareness of the message that is going to be communicated through integrated reports. Users have at their disposal a technique that could contribute to a better decoding of the message projected through integrated reports. Finally, regulators, specifically the IIRC, have at their disposal a technique for analyzing integrated reports and, consequently, evaluating the effectiveness of guidelines issued in driving preparers while drafting integrated reports.

The paper is organized as follows. Section 2 reviews studies on IR and IT and provides the theoretical framework on vocabularies for analyzing integrated reports. Section 3 explains the research design and the methodology applied. Section 4 presents the findings of the study and Section 5 provides a discussion of the results. Finally, Section 6 highlights some concluding remarks and new research directions.

2. Theoretical framing

In this section, we analyze accounting studies, which have highlighted the potential of IR for understanding and managing the increasing complexity of the business environment, as well as its relationship with the concept of IT (Section 2.1). By drawing on this research and on organizational studies, which emphasized a vocabulary approach to language, we argue that this approach may help to understand the structure of meaning embedded in integrated reports (Section 2.2).

2.1 Complexity, information and meanings: explaining the relationship between integrated reporting and integrated thinking

Organizations face more and more difficult challenges because of the increasing complexity of their environment. A weak management of this complexity could compromise their survival due to their loss of legitimacy within the environment in which they operate (Child and Rodrigues, 2011). In line with this perspective, organizations are increasingly called upon to implement, in compliance with the principle of requisite variety (Lengnick-Hall and Sanders, 1997; Poulis and Poulis, 2016), managerial innovations that allow them to understand and manage the complexities they face.

The main effect of the increasing complexity is that organizations' decisions that are made to achieve specific objectives affect different dimensions of their performance. This implies the need to simultaneously consider different performance dimensions during the decision-making processes carried out within an organization, to be able to indirectly consider all the subjects and objects involved in (and affected by) those decision-making processes. This need led to the appearance and diffusion of an integrated perspective within management practices that would make it possible to overcome the silo perspective underlying the systemic approach to organizations and to move toward IT (Lengnick-Hall and Sanders, 1997; Oliver *et al.*, 2016).

According to Checkland (1981), it is possible to classify system thinking as being hard or soft. The former focuses on the identification of resources and technologies that are instrumental for

pursuing a given objective that is stable and unchangeable for a defined system. Differently, soft system thinking “is more aligned with environments that are complex, messy, ill-defined and not independent of people and their individual value systems” (Oliver *et al.*, 2016: 232). Therefore, the hard perspective is much more in line with the Taylorist perspective of management, while soft system thinking is closer to the IT managerial perspective.

Coherent with the development of system thinking, during the past decades reporting systems have experienced a broad process of innovation, which has led to provide non-financial information together with classic financial information (Adams, 2004; Dossi and Patelli, 2010; Hoque, 2014; Kaplan and Norton, 1992). This innovation initially occurred through the preparation of specific reports, i.e. social and environmental reports (Adams, 2004; Contrafatto and Burns, 2013; Gray, 2002; Monfardini *et al.*, 2013) and intellectual capital statements (Guthrie and Dumay, 2015; Mouritsen, 1998). Subsequently, starting in 2010, the IIRC triggered a report integration process by developing the idea of an integrated report. In this report, the integration of the different dimensions of performance has not been conceived of as the simple sum of the different types of performance information available within an organization (Busco *et al.*, 2013), but as the result of a specific process aiming at providing:

[...] a concise communication about how an organization’s strategy, governance, performance and prospects, in the context of its external environment, lead to the creation of value over the short, medium and long term (International Integrated Reporting Council, 2013: 7).

Therefore, the main objective of IR is to bring together, adopting a holistic perspective, a series of information types regarding different business dimensions. IR stands in the middle between hard and soft system thinking, allowing to make visible the connection between performance measurement systems used for managing organizational resources and managerial skills while managing those resources in an integrated way. Churet and Eccles (2014, p. 8) sustain that “integrated reporting is only the tip of the iceberg. It is the visible part of what is happening below the surface, namely, ‘IT’ and ‘integrated decision-making’.” Accordingly, the International Integrated Reporting Council (2013, p. 2) has stated that the provision of a good integrated report depends on the level of IT as it is the active consideration by an organization of the relationships between its various operating and functional units and the capitals that the organization uses or affects diffused within an organization. The more that integrated thinking is embedded into an organization’s activities, the more naturally will the connectivity of information flow into management reporting, analysis and decision-making.” Al-Htaybat and von Alberti-Alhtaybat (2018) have found that organizations used to handling uncertainty and disruption through an integrated approach to their business are much more able to move toward the adoption of IR, being its natural extension. In addition, through their study, these scholars have highlighted that “commitment from the top is of utmost importance if IT, IR and long-term value creation and sustainability are to take place” (Al-Htaybat and von Alberti-Alhtaybat, 2018, p. 1437).

At the same time, integrated reports could contribute to increasing the effectiveness of IT within an organization by making the objects affected by IT material and developing a common understanding of this concept within that organization. In the end, between IT and integrated reports, there could be a virtual cycle (Figure 1). On the one hand, the existence of strong and effective IT within an organization could contribute to the drafting of integrated reports of higher quality. On the other hand, integrated reports could help give meaning to the concept of IT and to building a common understanding, not only for organizations’ external stakeholders but also for their internal stakeholders, such as managers and employees (International Integrated Reporting Council, 2016). Black Sun (2012) found that IR is posing different challenges to organizations, but at the same time, it is contributing to



Source: Our elaboration

Figure 1.
Relationship between
integrated thinking
and integrated report

drive positive changes within organizations. In particular, among others, what is relevant for this paper is that the practice of IR is contributing to connecting teams across organizations' departments, especially because of the increased visibility of (and managers' engagement with) organizational activities and the holistic value creation process.

According to the [World Intellectual Capital Initiative \(2013, p. 1\)](#):

[. . .] the experience has shown that integrated reporting in an organization further helps to embed integrated thinking. Therefore, the processes of Integrated Reporting and integrated thinking are mutually reinforcing. The organization's ability to maintain a big picture view, connect time horizons and develop a strategy for consistent messaging fosters a connected mindset and informs report content.

As a consequence, integrated reports are moving from being a promising accounting innovation, discussed more than applied ([Gibassier et al., 2018](#)), to a powerful practice, able to make IT more effective within organizations by making visible, the capitals which are mobilized to create value and how those capitals are connected ([Busco et al., 2017](#)).

Some scholars have investigated the concept of IT with reference to IR. [Guthrie et al. \(2017\)](#) have shown how the implementation process of IR in some public sector organizations has stimulated an internalization process of new managerial concepts and perspectives, resulting in the adoption of IT. However, in other cases, the vagueness of the concepts (such as IT) included in the IR framework, has been considered as an important obstacle to its adoption ([Dumay et al., 2017](#)). Vagueness in the IT concept mainly comes from "the absence of clear precedents in reporting context" ([Feng et al., 2017: 334](#)). Focusing on the measure of management's commitment and effectiveness toward integrating financial and extra-financial aspects, [Busco et al. \(2019\)](#) have defined and classified the possible levels of integration among IT and IR and the related drivers. In particular, four different levels of integration (i.e. holistic, integrated, conservative and minimalist) have been identified and higher levels of integration were found to be due to the greater size and leverage of companies, bigger board size and meetings, and to companies' more sensitive and higher environmental performance of their industries. Furthermore, the implementation of IR and its related concepts could be differently perceived by people operating at various organizational levels or it could produce different effects according to the "health" of an organization.

Considering the vagueness of the concept of IT, as well as the contribution that IR could have to make IT more and more effective within organizations, the understanding of whether and how the concept of IT is represented in integrated reports becomes more

relevant. Like other control and reporting systems, integrated reports are gears that turn management principles into everyday practices. They can contribute to establishing the recipes and rituals that govern the work of managers (Hamel, 2006). In line with this idea, IR is a signifying practice able to give meaning to the concept of IT and, at the same time, to produce a common understanding of it within an organization (Busco *et al.*, 2018). IR, and especially its final output (i.e. integrated reports), makes visible and material the different capitals that an organization acts upon throughout its business process and draws the images of the links existing among those capitals during a certain time period. Both capitals and links are the objects to which IT refers to acquire a meaning. Therefore, both internal and external stakeholders find in integrated reports the images of the capitals used by an organization (in terms of *which* capitals and *how* they are composed), and how they are jointly used and their evolution over time. According to Weick (1995, p. 4), this production of images is important because “organizations with access to more varied images will engage in sensemaking that is more adaptive than will organizations with more limited vocabularies.”

2.2 Vocabularies and integrated thinking

Following the linguistic turn in organizational research (Alvesson and Kärreman, 2000), a stream of literature has emphasized a vocabulary approach to language and meaning (Loewenstein *et al.*, 2012). According to this approach, vocabularies are defined as “systems of words, and the meaning of these words, used by collectives at different levels of analysis – groups, organizations, communities of practice, institutional fields-in communication, thought and action” (Loewenstein *et al.*, 2012, p. 45). Vocabularies are grounded in existing practices, which characterize these collectives and, at the same time, can be a source of new meanings. Indeed, they provide meaningful and shared categories, which guide thought, action and interaction between different groups of individuals.

Meaningful categories are structured around words or compound terms, (defined as conventional linguistic expressions – see Loewenstein and Ocasio, 2005), and the emergence of new categories can prompt new vocabularies which influence (and, in turn, are influenced by) how collectives make sense of (and act upon) complex organizing practices. However, despite their socially-agreed upon nature, categories (and their vocabularies) are not necessarily shared entirely within or across different collectives. Along these lines, written texts can be the expression of distinct vocabularies, which reflect diverse categories, and their analysis may allow for the identification of the specific “vocabulary use” by social actors (Jones and Livne-Tarandach, 2008).

In this context, words (and the categories they represent) gain relevance not as single and unconnected words, but as part of a specific vocabulary structure. Vocabulary structure relies on the combination of word frequency patterns (i.e. the meaning of more frequently used words), word-to-word relationships (i.e. how words relate to other words) and word-to-example relationships (i.e. how words relate to specific organizing practices in terms of objects, events, activities, ideas, etc., which make vocabularies concrete and understandable, see Loewenstein *et al.*, 2012, p. 45). In this vein, vocabulary structure can highlight the significance of a word, as well as what a category means (and its boundaries). The closely associated words which are used to invoke a category, as well as the properties of the examples to which these clusters of words have been applied, convey the meaning assigned to the category.

By extending these insights to the specific context of IR, we argue that the written texts of the integrated reports can be the expression of the specific “vocabulary use” by organizations, which reflects the category of IT. As suggested by recent accounting studies, IT is the result of a complex process of internalization (Guthrie *et al.*, 2017) and builds on

existing internal practices (Mcnally *et al.*, 2017) and cultures (Dumay and Dai, 2017). In this vein, we can consider IT as a meaningful category, which guides thought, action and interactions among organizational members. Consequently, IT can not only be operationalized in different ways as highlighted by recent literature (Dumay *et al.*, 2017; Feng *et al.*, 2017) but it can also influence IR, that is, the complex organizing practice organizations develop to face growing business complexity and through which IT takes form and concretely materializes.

In this perspective, by relying upon the idea of IR as a signifying practice (Busco *et al.*, 2018), we argue that it is through IR that the meaningful category of IT is projected. To understand the system of meanings underlying this category, we should focus on the vocabulary structures (Loewenstein *et al.*, 2012) that emerge in the integrated reports. This may allow understanding into how organizations make sense of IR practices and how, in turn, these practices convey the meanings of IT. Moreover, since a shared vocabulary “serves as a common ground on which members can draw to be understood” (Loewenstein *et al.*, 2012, p. 62), its comprehension can provide common ground to share and spread meanings also between organizations and their stakeholders (to whom integrated reports are also addressed) throughout organizations’ search for legitimation.

In the following sections, we show and discuss the findings of the analysis of the vocabulary structure reflecting the category of IT within a set of selected integrated reports, which serve as our empirical case. According to the IR framework and some policy position papers (Deloitte, 2015; International Integrated Reporting Council, 2016), IT is strictly related to the use and transformation of different kinds of capital and their interdependencies. In this paper, we argue that both capitals and their linkages can, thus, be considered as objects through which IT is represented in the reports. We focus, in particular, on the analysis of word frequency patterns and word-to-word relationships to identify such objects and their structure of meaning for the representation of the category of IT.

Before presenting the results of the analysis, the next section explains the overall research design.

3. Research design

The empirical material for the analysis consists of the integrated reports published by five organizations from 2012 to 2018, for a total amount of 34 reports (see Table A1 in the Appendix for details). The reports have been selected from the IIRC online database, section “recognized reports.” This database was chosen because it includes reports that have been recognized as leading practice by specific awarding processes or through benchmarking. Since our aim has been to explore the meanings of IT projected longitudinally through the integrated reports, we selected early reports that were awarded in 2013 and 2014 and then have been continuously published in the following years (independently from industry or type of organization). The publication of integrated reports was compulsory for some of the organizations selected, while for the others it was not. However, as the objective of this paper is to understand if and how IT materializes through the development of integrated reports, and not the content of that materialization, the compulsory or voluntary adoption of IR is not a bias of the findings of this paper. In addition, the integrated reports used in the analysis are of different lengths. Since our analysis is a longitudinal analysis that does not focus on individual reports, the difference in their length could lead the longer reports (and their content in terms of lemmas) to dominate over the shorter reports.

Our purpose was to provide neither generalized results regarding IR as a signifying practice for IT nor a complete overview of IR by early adopters. Rather, we have aimed to explore whether and how the structure of meaning embedded in the texts of integrated

reports could reveal the way in which IT can materialize and change over time. Given this objective, even if the reports could have been produced differently, such as internally or outsourced to a consultant, this is not relevant for the aims of the paper. Actually, we have focused on the existence of a relationship between IT and IR and its related materialization, rather than focusing on the effect caused on such relationship by a different integrated report production process. Previous research has already highlighted that analyzing the content and textual attributes of published integrated reports can help assess their compliance to the framework (Wild and van Staden, 2013), their changes compared to prior reporting practices (Stent and Dowler, 2015) and their level of readability (Du Toit, 2017; Caglio *et al.*, 2019) and sustainability disclosure (Montecalvo *et al.*, 2018; Zappettini and Unerman, 2016). Nevertheless, the ways in which vocabulary structure within the text of an integrated report can convey the meaning of IT is still an area that is open for contribution.

Reports were retrieved from the IIRC database or the organizations' websites and downloaded in PDF format. They were initially examined in their entirety to verify that they explicitly referred to the IR framework, as well as identify their structure in terms of the combination of texts, charts, images and figures. Then, PDF files were converted into TXT format for performing the analysis. Specifically, the paper combines textual analysis (Lebart *et al.*, 1998) and network text analysis (Carley, 1994; Diesner and Carley, 2005). They were used to identify the vocabulary structure of the selected integrated reports with specific reference to word frequency patterns and word-to-word relationships.

Textual analysis allows investigation of the use of words within texts and finds significant patterns of meaning. Thus, it was useful to our aims to identify specific concentrations of words that could be interpreted as different types of capital, through which the meanings of the category of IT could manifest itself, as well as their possible evolution over time. We carried out textual analysis by using the software T-Lab (Lancia, 2019). TXT files were uploaded to the software resulting in a global corpus of 19,009 words. This quantity of words was then processed. For instance, compound words (e.g. integrated report, risk management, etc.) and phrasal verbs were merged and all the stop-words in text (e.g. articles, prepositions, etc.) were discarded. From the remaining words, we selected and lemmatized the most significant words according to the TF-IDF measure (Salton, 1989). In so doing, the final corpus analyzed was composed of approximately 2,950 lemmas.

Textual analysis followed two main stages. In the first stage, we carried out a lexical correspondence analysis – LCA (Lebart and Salem, 1988), which is a statistical explorative technique that allows a reduction of the data's high dimensionality into fewer dimensions, with a loss of only a very small quantity of information (Greenacre, 1984; Lebart *et al.*, 1984; McEwan and Schlich, 1991). This technique is applied on contingency tables with size " $n \times m$ ", where " n " is the number of rows (words) and " m " the number of columns (text documents containing those words). In a dimensional perspective, each column (document) can be considered as a point in a n -dimensional space, (where each word is a coordinate in the space and the frequency of that word in a document is the document value respect to that coordinate), and can be compared to another column by confronting their rows' values (i.e. presence, absence and frequency of each word). The same approach can be used to compare two rows (words) considering their presence in the documents. Since high dimensional spaces are difficult to visualize and interpret, LCA creates a lower-dimensional space (called biplot when the space is bi-dimensional, see Gabriel, 1971) that preserves part of the original information in compared rows and columns. That is to say, the differences between two documents can be visualized by considering not all the n -dimensions representative of words, but only two dimensions created by the LCA. Specifically, to carry out the LCA, we used a bag-of-words method (Loughran and McDonald, 2016) to create a two-way contingency table, with reports of

the specific year of publication as columns and lemmas as rows. In this way, the LCA allowed us to visualize columns (years) and rows (lemmas) on a low dimensional space, to investigate the semantic orientation of the texts of the integrated reports.

In the second stage of the textual analysis, we conducted a thematic analysis of elementary contexts (Lancia, 2019). This analysis “creates a map of themes characterized by the co-occurrences of certain semantic traits” (Trobia and Lo Verde, 2017, p. 501). It considers elementary contexts (i.e. short parts of text, such as sentences of the entire corpus) that share similar words and it allows the development of a contingency table where each row is an elementary context and each column represents a lemma. Each cell of this table contains a value of presence or absence of the specific lemma in the elementary context. From this table, we performed a clustering procedure by using the bisecting K-means algorithm (Savaresi and Boley, 2001; Steinbach *et al.*, 2000). In so doing, our aim was to identify a map of the main themes depictive of different capitals.

Network text analysis is a text analysis method which “enables the extraction, analysis and concise representation of the complex network structure that can be represented in texts” (Diesner and Carley, 2005, p. 84. See also Carley and Kaufer, 1993). This analysis enables the interpretation of the meanings underlying texts by investigating the links between words and by constructing a network of the linked words. We relied upon a traditional proximity approach (Carley, 1993), but by using Markov chains based on co-occurrences to compute the probability of a word being the predecessor or successor of another word (Lancia, 2019). Specifically, we elaborated the networks by using the software T-Lab and then uploaded them to the software Cytoscape (Shannon *et al.*, 2003) for improving their visualization and interpretation. The analysis was first conducted to explore the links between words and, then, to explore links between the thematic clusters identified through the thematic analysis of elementary contexts. This analysis was useful for our aims to investigate whether and how capitals (through which IT manifests its meaning) are connected to each other within the text of the reports.

The combination of textual analysis and network text analysis allowed us to examine the vocabulary structure (i.e. word frequency patterns and word-to-word relationships), invoking the category of IT embedded in texts of integrated reports to explore how capitals and their linkages provide a meaning to IT. The results of this analysis are described in the next section.

4. Findings

4.1 Identifying latent meanings within integrated reports

The textual analysis resulted in a two-way contingency table (Table 1) with the years of report publication (i.e. from 2012 to 2018) as columns and lemmas as rows. Each cell of the contingency table contains the number of occurrences of a specific lemma in the integrated reports of a particular year. The analysis reveals a significant relationship between the two variables, with a chi-square statistics p -value lower than 0.05.

The LCA allowed us to further explore the relationship between lemmas and years in the contingency table by reducing the dimensions of analysis. Specifically, we obtained six dimensions that were able to explain the latent meaning of the corpus. As the first two dimensions together explain the most part of the total amount of information (i.e. approximately 70%), we represented our findings through a bi-dimensional factor space or biplot (Figure 2). On this space, points (i.e. lemmas and years) are distributed according to specific coordinates, which identify the positioning of each point with reference to the two dimensions (i.e. the X and Y axes).

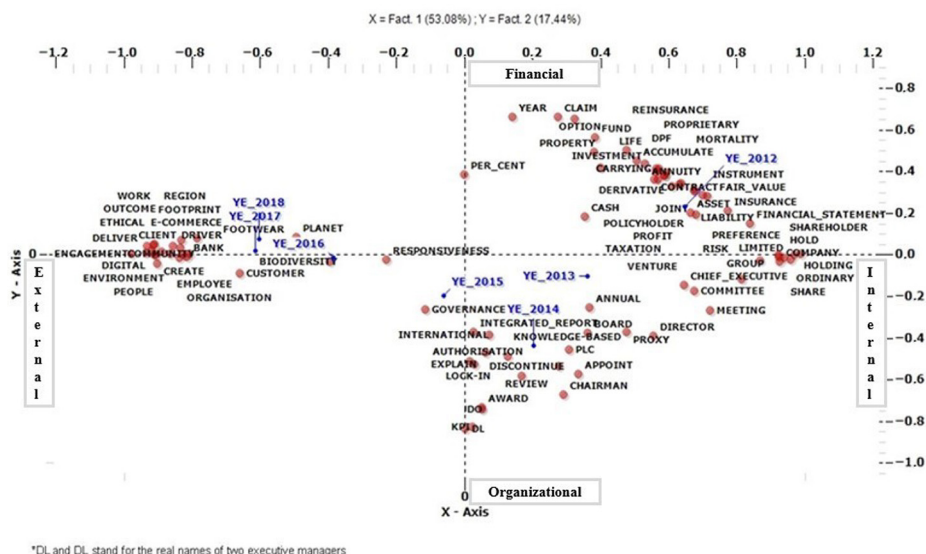
LEMMA	2012	2013	2014	2015	2016	2017	2018
A3	1	1	0	0	0	4	5
AA1000	9	10	0	1	17	3	2
ABBREVIATION	11	9	9	7	5	2	2
ABERDEEN	3	5	4	0	1	0	0
ABILITY	82	80	85	86	86	97	90
ABLE_TO	45	42	32	30	21	27	39
ABORIGINAL	1	1	1	3	8	7	14
ABRIDGE	28	40	27	1	4	0	0
ABSORB	7	2	1	0	3	2	1
ACADEMY	1	0	1	0	10	9	9
ACCELERATE	7	12	12	15	23	39	23
ACCEPT	30	36	17	7	9	11	9
ACCEPTABLE	21	17	7	7	7	9	11
ACCEPTANCE	8	13	7	6	5	5	3
ACCESS	83	73	47	80	84	92	111
ACCESSBANKING	3	6	2	0	0	0	0
ACCESSORY	7	8	8	5	12	9	10
ACCIDENT	10	2	0	0	1	3	0
ACCOMMODATION	2	1	1	0	0	2	9

Table 1.
Lemmas × years
occurrences
contingency table

Notes: Table 1 reports a part of the two-way contingency table with years of report publication (from 2012 to 2018) as columns and an example (in alphabetic order) of all lemmas (approximately 2,950) as rows. The numbers in each cell indicate the number of occurrences of a specific lemma in the reports of a specific year. Pearson’s Chi-squared test: X-squared = 116,575; df = 17,622; *p*-value < 2.2e-16
Source: Part of the total output

To make sense of the meanings underlying these two dimensions, we analyzed the contribution to inertia value, which is the percentage of information that a single point in the space explains about each of them (see Table A2 in the Appendix). In other words, the points with the highest contribution to inertia values help in the interpretation of the meaning that the two dimensions imply. In our case, we found that the lemmas with the highest contribution to inertia values for the X-axis depict a semantic opposition between an *internally-oriented perspective* (e.g. see lemmas such as “share”, “company”, “group,” etc.) and an *externally-oriented perspective* (e.g. see lemmas such as “client”, “customer”, “community”, etc.). As we can observe in Figure 2, the internally-oriented perspective takes shape on the right side of the X-axis, while the externally-oriented perspective on the left side. As for the Y-axis, the lemmas with the highest contribution to inertia values seem to depict a semantic opposition between an *organizational perspective* (e.g. see lemmas such as “authorization”, “chairman”, “director”– the downside of the Y-axis) and a *financial perspective* (e.g. see lemmas such as “liability”, “investment”, “fair value” – the upside of the Y-axis).

Points with opposite meaning with respect to a dimension can be found at the opposite side in the factor space. For instance, in this context, the lemmas “company” (at the extreme right side of the X-axis) and “engagement” (at the extreme left side of the X-axis) have opposite meanings with respect to the internal/external perspective. In addition, as the more the points are distant from the origin of axes, the more they are significant for the analysis, a year and a lemma close to each other and yet far from the origins have a strong relationship (i.e. the reports published in that year are highly associated with that lemma).



Notes: Figure 2 depicts the bi-dimensional factor space in which lemmas and years of the reports are points distributed on the space according to specific coordinates with reference to two dimensions of analysis (the X and Y axes). The X-axis illustrates the semantic opposition between an internally-oriented perspective and an externally-oriented perspective, while the Y-axis illustrates the opposition between an organizational perspective and a financial perspective

Source: Our elaboration from the empirical material

Figure 2.
Lemmas x years
biplot

As shown in [Figure 2](#), the years are distributed longitudinally on the factor space showing a shift in the lemmas (and their semantic orientation) included in the reports from 2012 to 2018. Early integrated reports (i.e. from 2012 to 2014) are positioned on the right side of the factor space and, thus, are strongly related to the internally-oriented perspective. Specifically, the first year of the reports (i.e. 2012) appears in the high-right quadrant and, thus, it is also related to the financial perspective (e.g. see lemmas such as “fair-value”, “asset”, “derivative”, “profit”, “cash”). Instead, in the low-right quadrant, we can find the years 2013 and 2014, whose main lemmas are still related to an internally-oriented perspective but more in terms of organizational issues (e.g. see lemmas such as “board”, “chairman”, “knowledge-based”). Moving toward the left side of the factor space, we find that more recent integrated reports (i.e. 2016, 2017, 2018) are strongly related to the externally-oriented perspective (e.g. see lemmas such as “community”, “engagement”, “environment”, “people”, “client”, “ethical”). Interestingly, the analysis suggests that there has been an evolution in the semantic orientation of the published integrated reports. This finding is also reinforced by the fact that the year 2015 lies near the origin of the axes and, thus, it is not particularly reflective of any dimension. This probably means that the 2015 integrated reports are related to all lemmas with different semantic meanings in a more balanced way, witnessing a slight shift over time.

To sum up, we have found that the concentration of words used within the integrated reports mainly reflects two opposite semantic meanings, which refer to an internal and an

external perspective. Importantly, the semantic orientation of the integrated reports seems to evolve over time, shifting from a focus on organizational and financial issues to a focus on more relational aspects between organizations and their environment.

Next, by analyzing the co-occurrences between words within specific elementary contexts, we show that these words give shape to a map of different capitals through which the meaning of IT materializes in the reports.

4.2 Revealing the capitals underlying the meaning of integrated thinking

The thematic analysis of elementary contexts allowed us to identify the main themes emerging from the corpus. Six main themes emerged and we found that each of them could be depictive of a different notion of capital (see [Table A3](#) in the [Appendix](#)). Although the various capitals emerging from the findings of this paper are named differently from the taxonomy provided by the [International Integrated Reporting Council \(2013\)](#), this is not intended to deny or criticize the IIRC's taxonomy. The six capitals identified through our analysis are indeed closely aligned with all the forms of capital that an organization uses or affects according to the IR framework. The different taxonomy provided in this paper is due to the choice of naming the different capitals according to the most representative words for each of the capitals disclosed by the analysis.

The first theme, which we labeled financial capital (FC1), contains lemmas with the highest chi-square [1] such as “asset”, “income”, “cost”, “liability”, “earning” and refers broadly to all the financial resources available to organizations through different forms of financing, as well as those generated through their activities. For instance, an elementary context with lemmas representative of this capital is:

Over the past year Bankmecu has continued its long-standing record of strong financial performance. Bankmecu's net profit after tax of \$27.02 million represents a return on assets of 0.99%. (Extract 1, Bankmecu, Annual Report, 2012, p. 2).

The second theme, labeled human capital (HC2), contains lemmas with the highest chi-square such as “award”, “employee”, “incentive”, “remuneration”, “reward” and includes the relationships between the organization and its members, with a particular focus on the mechanisms and the processes for enhancing competences and motivation. For instance, an elementary context with lemmas representative of this capital is:

Attract and retain quality employees. Employees are integral to delivering the group's strategies. Consequently remuneration structures are designed to create a high performance culture with the necessary balance between short and long-term objectives. Skills development, engagement processes and wellness support are designed to assist retention and maximise productivity. (Extract 2, Liberty Holdings Limited, Integrated Annual Report, 2012, p. 2).

The third theme, which we labeled social-relational capital (SRC3), contains lemmas with the highest chi-square such as “community”, “customer”, “biodiversity”, “satisfaction”, “education” and includes all the relationships an organization develops with its multiple stakeholders with a focus also on social and environmental issues. For instance, an elementary context with lemmas representative of this capital is:

We believe in adopting a broader measure of performance than economic growth alone. By advancing the economic, social, environmental and governance conditions in the communities where we operate we can offer a better kind of bank. (Extract 3, Bankmecu, Annual Report, 2012, p. 0).

The fourth theme, labeled process capital (PC4), contains lemmas with the highest chi-square such as “report”, “risk”, “governance”, “compliance”, “information” and refers in broad terms to all the techniques and procedures that allow the enabling, managing and

monitoring of organizational activities. For instance, an elementary context with lemmas representative of this capital is:

We regularly assess and enhance our risk management framework to ensure that it is fit-for-purpose and that we have adequate capacity to manage risk in unpredictable operating environments. (Extract 4, Standard Bank, Annual integrated report, 2016, p. 71).

The fifth theme, which we labeled organizational capital (OC5), contains lemmas with the highest chi-square such as “director”, “board”, “committee”, “chairman”, “chief” and includes all the roles and functions whose coordination and interaction foster internal processes. For instance, an elementary context with lemmas representative of this capital is:

The performance of the board and its committees is evaluated periodically against their respective mandates and the results are collated by independent assurance providers. Feedback is provided to the directors’ affairs committee and thereafter to the board. (Extract 5, Liberty Holdings Limited, Integrated report, 2017, p. 20).

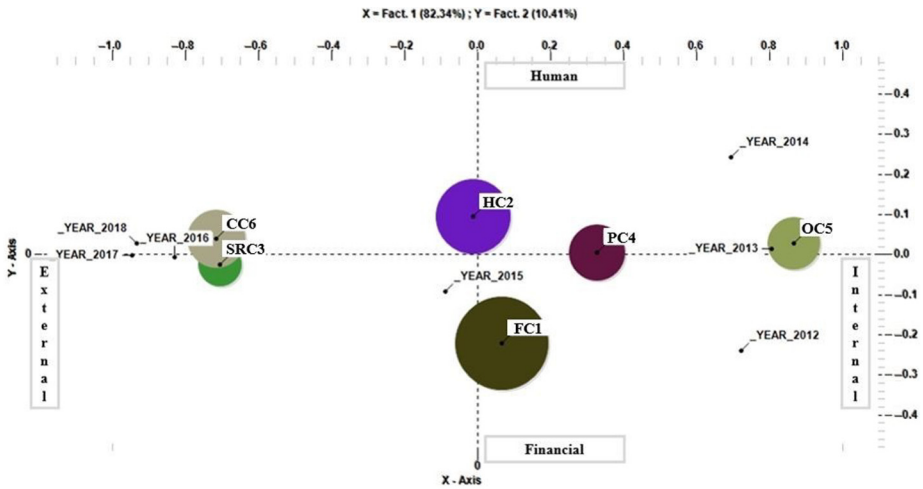
The sixth theme, labeled commercial capital (CC6), contains lemmas with the highest chi-square such as “store”, “client”, “brand”, “product”, “service” and refers broadly to all the relationships between an organization and its external market. For instance, an elementary context with lemmas representative of this capital is:

Our focus is to consistently create excellent client experiences, by understanding our clients and by offering the products, services and solutions they need (Extract 6, Standard Bank Group, Annual Integrated Report, 2016, p. 12).

Figure 3 presents the main findings of the LCA, which has been conducted on another contingency table with years (as rows) and themes (as columns) to observe how the six notions of capitals are positioned on a factor space. In this case, we obtained five dimensions to explain the meaning of the corpus. However, as the first two dimensions together explain approximately 92% of the total information, we still represented our findings through a biplot.

By analyzing the contribution to inertia values, we were able to interpret the meaning that the two dimensions imply. Specifically, we found that the highest contribution to inertia values for the X-axis refer to both the organizational capital (OC5) theme and the commercial capital (CC6) theme. Therefore, considering themes (rather than lemmas), the X-axis still seems to depict the same semantic opposition of the lemmas per years biplot (i.e. an internally-oriented perspective, on the right side of the X-axis, and an externally-oriented perspective, on the left side of the X-axis). Conversely, the analysis of the contribution to inertia values (as well as of other quality measures) for the Y-axis revealed the prominence of the financial capital (FC1) theme and, more outdistanced, of the human capital (HC2) theme. Thus, if we consider themes (rather than lemmas), the Y-axis seems to depict a semantic opposition similar to that of lemmas per years biplot. We find a financial perspective, on the downside of the Y-axis, and an organizational perspective, but more human-oriented (very close to the idea of human capital provided above), on the upside of the Y-axis.

As shown in Figure 3 by the size of bubbles, independently from the years of publication, financial capital (FC1) is the most discussed theme within the all reports (25.79% is the percentage of elementary contexts referring to this theme). Then this theme is followed by human capital (HC2- 20.23%), commercial capital (CC6- 15.01%), process capital (PC4- 14.89%), organizational capital (OC5- 13.22%) and, finally, by social-relational capital (SRC3- 10.87%). If we focus on the position of years and themes, we find that the first triennium (from 2012 to 2014) and the last triennium (from 2016 to 2018) are semantically in opposition along the internal/external dimension. Early integrated reports (i.e. 2012, 2013



Notes: Figure 3 depicts the bi-dimensional factor space in which themes (FC1-financial capital, HC2-human capital, SRC3-social-relational capital, PC4-process capital, OC5-organizational capital, CC6-commercial capital) and years of the reports are distributed on the space with reference to two dimensions of analysis (the X and Y axes). The X-axis illustrates the semantic opposition between an internally-oriented perspective and an externally-oriented perspective, while the Y-axis illustrates the opposition between a financial perspective and a human perspective

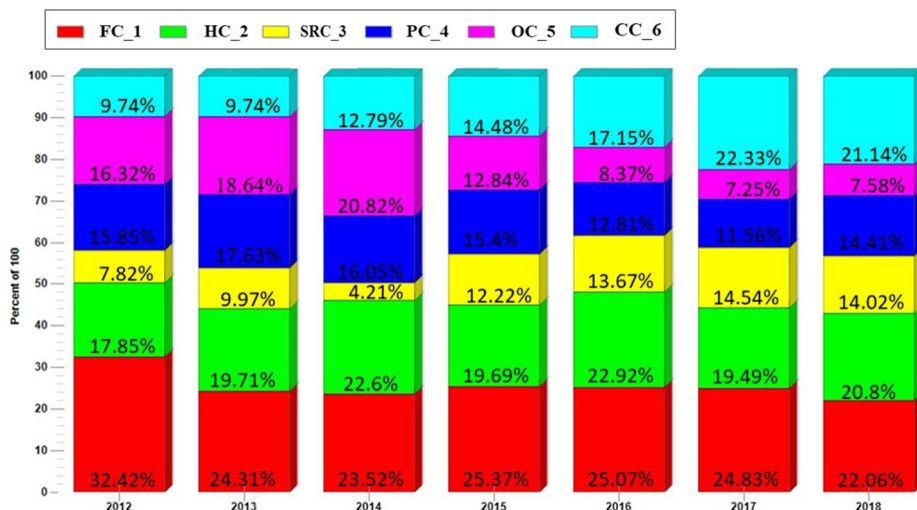
Source: Our elaboration from the empirical material

Figure 3.
Years × themes
biplot

and 2014) lie in a similar position with respect to the X-axis, confirming their semantic orientation toward an internally-oriented perspective. The first year (i.e. 2012) is in the lower-right quadrant, thus corroborating the semantic prevalence of the financial perspective in the reports of that year. From 2013, there is a growing opening up toward the human perspective. Indeed, 2014 is the highest point with respect to the Y-axis. Also, 2013 is the closest point to the organizational capital theme (OC5), which is shifted toward the internally-oriented polarity of the factor space. Moving toward the left side of the X-axis (i.e. toward an externally-oriented perspective), we can find reports published in 2016, 2017 and 2018. Interestingly, the points referring to the last triennium are very close to each other compared to the closeness between the points referring to the first triennium. This suggests a pattern of strengthening toward the two themes, which almost overlap at the externally-oriented polarity of the space, that is social-relational capital (SRC3) and commercial capital (CC6). The closeness of the year 2015 to the origins of axes still corroborates the idea of a slight shift over time.

To better understand the dynamic composition of the integrated reports, in Figure 4 we can observe the weight of all themes (i.e. capitals) for each year of the reports. As shown in the figure, the financial capital (FC1) has the highest weight in each year. Moreover, the changing pathway in the compositions of the reports is confirmed by the increasing weight of the capitals, which relate more to the relations between the organizations and their environment (i.e. the commercial capital – CC6 and the social-relational capital – SRC3).

Overall, we have found that IT has been signified and projected over time through the integrated reports in different ways. The dynamic nature of the meaning of IT depends on



Notes: Figure 4 depicts the percentage weight of all themes (FC1-financial capital, HC2-human capital, SRC3-social-relational capital, PC4-process capital, OC5-organizational capital, CC6-commercial capital) for each year of the reports (from 2012 to 2018)
Source: Our elaboration from the empirical material

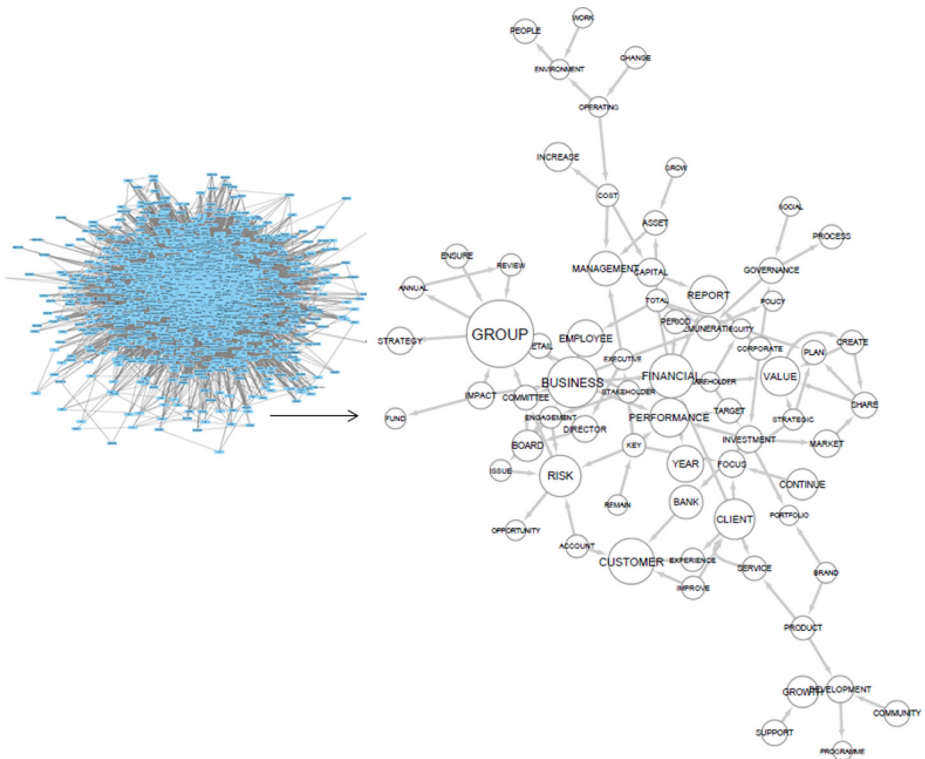
Figure 4.
Weights of the themes for each year

the changing weight of the multiple capitals represented within the reports during the years. While the map of the main themes emerging from the reports allowed us to find the capitals through which the category of IT materializes, next we focus on the linkages among words and among capitals to find different paths of their interdependencies, which shows that the meaning of IT is continuously unfolding.

4.3 Constructing networks of words and capitals

To see how lemmas are related to each other, we have constructed networks of all the lemmas within the corpus with reference to the extreme years (i.e. 2012/2013 and 2017/2018) of our time period. This choice is due to the fact that, from our previous analysis, we found that these two intervals highlight the main changes. Figure 5 shows the network originating from the data for the period 2017/2018. In this visual representation, all the nodes represent the lemmas in the corpus and an edge between two nodes is given by the probability of a lemma to be predecessor or successor of another lemma. A node with several edges is representative of a lemma having connections with diversified words.

To interpret the network of words, we can rely upon some measures of semantic connectivity that is density, betweenness centrality (BC) and conductivity. These measures are proxies of the role of the node in the network and, thus, of the words in the text. The density represents the total number of in-and-out links that directly connect a node (i.e. the lemma in our case) to the other nodes. The higher the density of a node is, the higher the level of prominence of this node is in the network. Thus, in our case, this measure expresses how much a lemma is “spread” through the text. Then, we can find the betweenness centrality and the conductivity. These are two main measures of centrality which allow the identification of the most influential nodes in connecting a network’s community of nodes.



Notes: Figure 5 represents, in the left-hand side, the network of all the lemmas with reference period 2017-2018. In the right-hand side of the figure, there is an enlarged and simplified (i.e. showing only the first 100 links between lemmas) detail of the network. The nodes are the lemmas. The edge between two nodes is given by the probability of a lemma to be predecessor or successor of another lemma

Source: our elaboration from the empirical material

Figure 5.
Network of words for
the period 2017-2018

Specifically, the betweenness centrality shows how often a node appears on the shortest path between two other random nodes in the network. The conductivity is the number of outgoing edges of a specific node, multiplied by the number of incoming edges for the same node. Thus, in our case, a lemma with high measures of centrality can be considered to be a connector, which binds other lemmas together.

Tables 2 and 3 show the measures of semantic connectivity of the first 30 lemmas with the highest measures of betweenness centrality, respectively, for the periods 2012–2013 and 2017-2018. This means that these lemmas are the most influential in connecting networks of lemmas.

Amongst the lemmas with the highest betweenness centrality listed in Tables 2 and 3, we can find lemmas which are the most representative of the identified themes (i.e. the capitals) in terms of both statistical significance (i.e. lemmas with the highest chi-square – see Table A3 in the Appendix) and conventional common sense (e.g. the lemma “employee” is conventionally referred to the notion of human capital).

Lemma	BC	Density	Dens (%)	Conductivity	Cond (%)
Group	0.09572826	1,705	0.009366	724,404	0.069044
Year	0.03627484	946	0.005197	223,153	0.021269
Business	0.03188644	1,031	0.005663	265,608	0.025316
Risk	0.02856008	951	0.005224	226,028	0.021543
Management	0.02047773	810	0.004449	163,241	0.015559
Report	0.01878152	778	0.004274	148,405	0.014145
Customer	0.01868082	791	0.004345	155,160	0.014789
Continue	0.01845473	774	0.004252	148,544	0.014158
Director	0.01838991	594	0.003263	85,808	0.008178
Share	0.01749347	649	0.003565	103,578	0.009872
Asset	0.01669194	717	0.003939	128,450	0.012243
Increase	0.01612833	683	0.003752	116,592	0.011113
Board	0.01605509	673	0.003697	112,042	0.010679
Employee	0.01593476	689	0.003785	118,674	0.011311
Market	0.0156223	664	0.003647	109,999	0.010484
Value	0.01491373	707	0.003884	124,260	0.011843
Performance	0.01358623	666	0.003658	110,664	0.010548
Investment	0.01314074	660	0.003625	107,136	0.010211
Company	0.01285571	609	0.003345	92,610	0.008827
South Africa	0.01204871	543	0.002983	73,530	0.007008
Total	0.01147828	542	0.002977	72,417	0.006902
Committee	0.01111491	474	0.002604	50,840	0.004846
Shareholder	0.01096146	541	0.002972	72,664	0.006926
Bank	0.01092353	499	0.002741	62,178	0.005926
Fund	0.01070355	553	0.003038	74,382	0.007089
Capital	0.01045699	563	0.003093	79,242	0.007553
Financial	0.0094447	605	0.003323	63,784	0.006079
Growth	0.00911503	535	0.002939	71,250	0.006791
Change	0.00893291	527	0.002895	69,360	0.006611
Development	0.00887658	520	0.002856	65,296	0.006223

Table 2.
Lemmas with the
highest measures of
betweenness
centrality for
2012-2013

Notes: Table 2 reports the first 30 lemmas with the highest measures of betweenness centrality (BC) for the period 2012-2013. For each lemma, the table also highlights the density of the node and the conductivity, both in absolute and percentage terms

Source: Part of the total output

Interestingly, we can notice that the lemmas representative of the human capital (HC2) such as “employee,” and of more relational capitals (i.e. commercial capital – CC6 and social relational capital – SRC3) such as “customer”, “client”, “community” have higher measures of betweenness centrality, density (in per cent) and conductivity (in per cent) in the years 2017/2018 than in 2012/2013. Some of these lemmas (such as “client” and “community”) are not even among the first 30 lemmas with the highest betweenness centrality in 2012/2013. This means that the prominence of these words in the past years of the time period increases and they become more influential in connecting networks of words. This corroborates the findings of the evolution of the semantic orientation of the integrated reports, which shift from being managerial-centric (in the early years) to be more social and relational-centric (in the past years).

To investigate the use of these lemmas (and their links) in the text in more depth, we extrapolated from the network only the links referring to specific lemmas. For instance, if we analyze the lemma “client” (which is highly representative of the commercial capital – CC6), we can consider all the lemmas to which it is connected. Figure 6 shows all these

Lemma	BC	Density	Dens (%)	Conductivity	Cond (%)
Group	0.06841	1,108	0.008101	305,472	0.047638
Business	0.037151	875	0.006397	190,650	0.029731
Customer	0.033057	792	0.00579	154,967	0.024167
Year	0.030549	712	0.005205	126,540	0.019734
Employee	0.024496	666	0.004869	110,633	0.017253
Increase	0.021507	611	0.004467	93,318	0.014553
Client	0.021355	669	0.004891	111,078	0.017322
Risk	0.020662	623	0.004555	96,850	0.015103
Financial	0.019382	624	0.004562	68,444	0.010674
Bank	0.018051	572	0.004182	81,760	0.01275
Report	0.01736	551	0.004028	75,144	0.011719
Performance	0.016598	569	0.00416	80,758	0.012594
Management	0.015844	553	0.004043	75,696	0.011805
Value	0.015422	526	0.003846	69,153	0.010784
Support	0.015177	561	0.004101	78,678	0.01227
Continue	0.014986	543	0.00397	73,656	0.011486
Market	0.01451	511	0.003736	65,224	0.010172
Growth	0.013895	515	0.003765	64,666	0.010084
Board	0.013231	482	0.003524	57,720	0.009001
South_africa	0.012683	435	0.00318	47,234	0.007366
Store	0.01248	372	0.00272	34,592	0.005395
Community	0.011706	431	0.003151	46,428	0.00724
Development	0.011691	465	0.0034	52,796	0.008233
Impact	0.011224	450	0.00329	50,621	0.007894
Investment	0.011098	444	0.003246	49,268	0.007683
Asset	0.01004	374	0.002734	34,960	0.005452
Million	0.009789	367	0.002683	31,510	0.004914
Focus	0.009436	445	0.003253	49,266	0.007683
Ensure	0.00913	419	0.003063	43,878	0.006843
People	0.009125	402	0.002939	40,337	0.00629

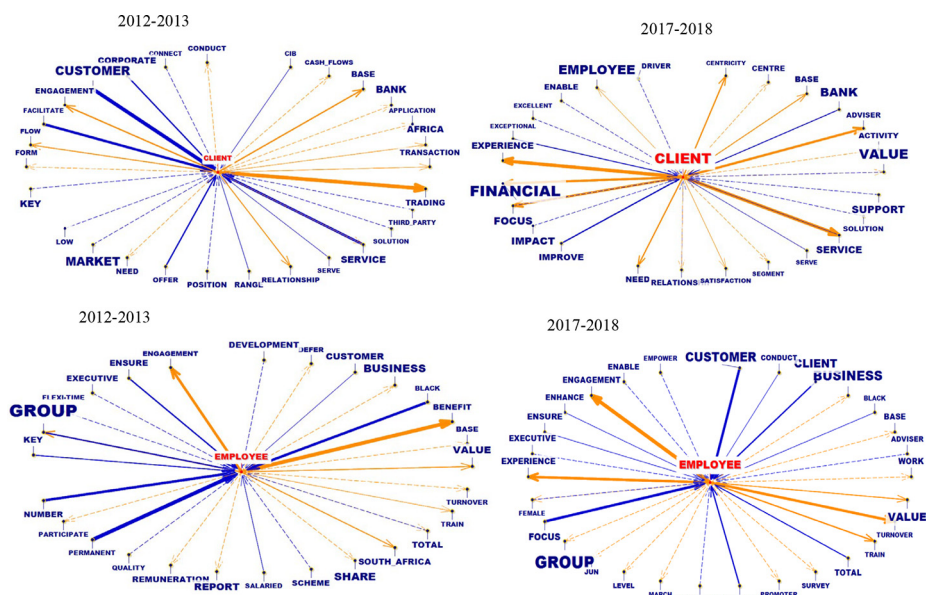
Table 3.
Lemmas with the highest measures of betweenness centrality for 2017-2018

Notes: Table 3 reports the first 30 lemmas with the highest measures of betweenness centrality (BC) for the period 2017-2018. For each lemma, the table also highlights the density of the node and the conductivity, both in absolute and percentage terms
Source: Part of the total output

connections. The arrows (and their direction) represent all the incoming and outgoing edges. The thickness of the arrows is proportional to the strength of the edge. We can see that in the period 2012-2013 the strongest edges are between “client” and the lemmas “trading” (incoming edge) and “facilitate” (outgoing edge). Thus, these two lemmas have the highest probability of being close to the lemma “client” in segments of text. Interestingly, “trading” is a highly representative lemma of the financial capital (FC1). An example is provided by the following extract:

Growth in trading revenue depends largely on trading volumes and how volatility affects trading spreads. The group’s trading revenue is mainly a function of client trading volumes and the margin between bid and offer prices. (Extract 7, Standard Bank Group, Annual Integrated Report, 2013, p. 64).

Instead, in the period 2017–2018, the strongest edges are between “client” and the lemmas “experience” (incoming edge) and “improve” (outgoing edge), as highlighted in the following extracts:



Notes: Figure 6 shows all lemmas connected with the lemma “client” (at the top of the figure) and the lemma “employee” (at the bottom of the figure) in the periods 2012-2013 and 2017-2018. The strongest incoming and outgoing edges are those represented by the thickest arrows
Source: Our elaboration from the empirical material

Figure 6. Examples of words’ one-to-one connections

As we become more digitised and integrated, we are better able to understand our client’s financial services needs and match them with personalised experiences and solutions. (Extract 8, Standard Bank Group, Annual Integrated Report, 2018, p. 38).

As for the lemma “employee” (which is highly representative of the human capital – HC2), for instance, we can see that in 2012-2013, the strongest edges are with the lemmas “benefit” (incoming edge) and “permanent” (outgoing edge), while in 2017-2018 with “engagement” (incoming edge) and “customer” (outgoing edge). Interestingly, “customer” is a highly representative lemma of the social-relational capital (SRC3). Examples are provided in the following extracts:

In 2012, the employee wellbeing programme was extended to Liberty entities throughout Africa extending access to benefits to approximately 600 employees. (Extract 9, Liberty Holdings Limited, Integrated Annual Report, 2012, p. 113).

Our employees provide us with many of our core competencies, including delivering excellent customer service, innovative product development, strong investment management skills and excellent balance sheet management. Liberty promotes a high-performance, customer centric culture to drive innovation and execution of strategy to create value for stakeholders. (Extract 10, Liberty Holdings Limited, Integrated Report, 2017, p. 55).

Therefore, networks of words allowed us to find that the meanings emerging from the reports can be dynamic in nature, as the links between words change over time, reflecting the changing semantic orientation of the reports. Specifically, we notice that changing links between words could be depictive of changing links between different capitals. To go deeper to this insight, we

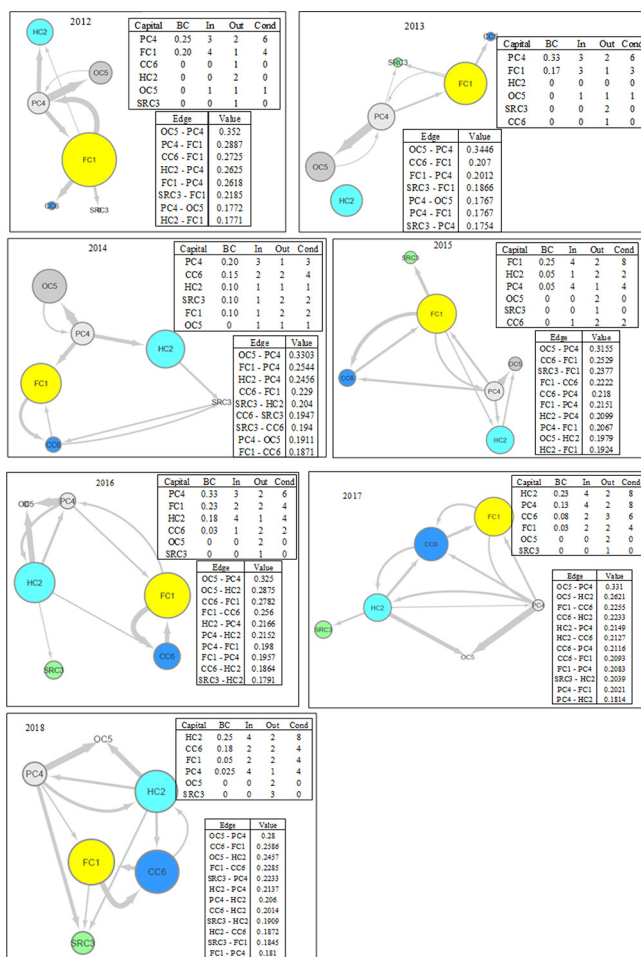
constructed networks between the six identified themes (i.e. the capitals) for each year from 2012 to 2018. Figure 7 shows the visual representation of these networks [2]. The nodes represent the themes (composed by elementary contexts). The size of each node depends on the weight of the capital in the reports of each specific year (as highlighted in the previous Figure 4), while the thickness of the arrows (and their direction) depends on the strength (and the direction) of the edge. The edge between two nodes is given by the sequence of elementary contexts related to different themes, that is, the probability that elementary contexts related to a specific theme are predecessor or successor of elementary contexts related to another theme (see the probability for each edge to occur in Figure 7).

By analyzing the measures of semantic connectivity (Figure 7), we can see different evolutionary patterns in the linkages between capitals over time. In 2012, while financial capital (FC1) is the main node in terms of size (i.e. the weight of the capital in the reports, see also Figure 4), process capital (PC4) is the node with the highest measures of connectivity. All the other capitals appear less relevant inside the net. Thus, this means that, whereas the content of 2012 integrated reports is more focused on financial issues, it is process capital (PC4) that represents the main connector (i.e. the process capital tends to bind the highest number of capitals together), followed by financial capital (FC1).

In the following four years, the situation is quite similar for financial capital (FC1) and process capital (PC4), which remain the most important nodes, only with some slight changes in 2014 (when the betweenness centrality of FC1 decreases). Interestingly, starting from 2014 the measures of connectivity of commercial capital (CC6), social-relational capital (SRC3) and human capital (HC2) begin to increase. This suggests that it is more likely that these capitals interact within the network.

However, we can find a sharp change in the last two years of the analysis (i.e. 2017 and 2018), when human capital (HC2) and commercial capital (CC6) become the main connectors. This confirms what resulted from the networks of words, which showed the growing prominence and connectivity of the main lemmas related to these capitals. This means that over time, human capital (HC2) and commercial capital (CC6) (and the words representative of these capitals) gained relevance as information that was able to connect information also related to other dimensions. In other words, while acting as connectors, these themes could carry larger amounts of information. Moreover, the probability for the edges between these two capitals to occur increases during this period, witnessing an increasing strength in their link. We can also notice that, while in 2012 social-relational capital (SRC3) is more connected with financial capital (FC1), with a probability value of 0.2185, in the following years there is a growing level of connectivity with other capitals as well, among which, above all, human capital (HC2). These findings confirm the shift in the semantic orientation of the reports toward an external perspective, also through a focus on human capital.

Overall, during the entire period, the strongest edges (in terms of the probability value for each edge) are generally those which link process capital (PC4) and organizational capital (OC5), whose borders can be typically overlapping and those which link financial capital (FC1) and commercial capital (CC6). However, over time, we can see an increase in the number of edges between all capitals (i.e. from 8 in 2012 to 12 in 2018), as well as the constant growth of the total level of conductivity (i.e. from 11 in 2012 to 20 in 2018). Thus, there has been a general growth in the connections between capitals. This growth is also confirmed if we calculate the clustering coefficient (i.e. a measure of the degree to which nodes tend to cluster together), which increases shifting from 0.25 in 2012 to 0.47 in 2018. These findings suggest that over time all capitals started to be bundled together, therefore, being more interdependent.



Notes: Figure 7 shows the networks representing the relationships between the six identified capitals (FC1-financial capital, HC2-human capital, SRC3-social-relational capital, PC4-process capital, OC5-organizational capital, CC6-commercial capital) for each year of the reports. Tables in the figure identify the probability for each edge to occur as well as, for each node of the network, the betweenness centrality (BC), the number of incoming (In) and outgoing (Out) edges, and the conductivity (Cond, i.e. the number of outgoing edges multiplied by the number of incoming edges)

Source: Our elaboration from the empirical material

Figure 7. Networks of capitals for each year

To sum up, we found that within the integrated reports the linkages between multiple capitals have changed over time, as well as the strength of these linkages. In so doing, the meaning of IT has taken different forms, which affect the semantic orientation of the reports.

5. Discussion

The analysis conducted on the selected integrated reports from 2012 to 2018 reveals the complex and dynamic nature of the concept of IT and its relationship with IR. By drawing on a vocabulary approach to language (Loewenstein *et al.*, 2012), our findings show that we can investigate more deeply the concept of IT by looking at the system of words, which outlines IT, as deemed as a category of meanings.

As argued by Loewenstein *et al.* (2012), words and their relationships gain relevance as part of a vocabulary structure, which identifies the specific “vocabulary use” by groups of individuals. This vocabulary reflects meaningful categories, which are grounded in existing practices characterizing these groups. In the context of IR, previous research has highlighted that IT originates from internal processes and cultures (Dumay and Dai, 2017; McNally *et al.*, 2017) and, thus, affects the flows of information, which result in integrated reports. We add to these studies by showing that this information is provided according to the specific “vocabulary use” by organizations, which makes the category of IT visible and material.

Specifically, in our case vocabulary use refers to the different capitals (and their linkages) through which IT is represented. As widely acknowledged (International Integrated Reporting Council, 2016; Chartered Institute of Management Accountants, 2017; Deloitte, 2015), thinking about multiple capitals and their interdependencies is indeed a form of IT, which takes shape according to the ways in which capitals are mobilized to create value (Busco *et al.*, 2017). In our analysis, we corroborate these insights by investigating word frequency patterns and word-to-word relationships, which allowed us to find six main themes depictive of different kinds of capital (i.e. financial, human, social-relational, process, organizational and commercial) within the reports. In addition, it is important to highlight that the representation of the different capitals could change according to the process of production of integrated reports which can be, for instance, internally produced or outsourced to a consultant. In this latter case, it would be important to investigate the interactions between consultants and companies to understand and analyze the level of decoupling that could be possible between the concept of IT reported in integrated reports and that existing within the organization.

Moreover, our findings show the potential dynamic nature of the meaning of IT due to the evolutionary paths of the weights of each capital within the integrated reports. In particular, in our case, we show the increasing weight of the more “relational” capitals throughout the period 2012-2018 (Figure 4). This dynamic nature also originates from the ongoing changes in the linkages between capitals, which are related to each other differently over time. For instance, in the case of the selected reports, through network text analysis, we found that in early reports the meanings of IT are conveyed mainly through financial capital and process capital (which are those capitals that most bind together all the information – see Figure 7). Instead, in the following years, all the capitals seem more interdependent, thus conveying the meaning of IT through wider dimensions of performance.

These findings are in line with the idea that the capitals used and transformed by organizations can change over time (Barnabè *et al.*, 2019). Notwithstanding, we also add to this idea by highlighting that the changing prevalence of the capitals, as well as their linkages, mirror changing semantic orientations emerging from the reports. As suggested by Loewenstein *et al.* (2012), shared categories do not only arise from existing practices but also guide thought and action, influencing how individuals make sense of (and act upon)

organizing practices. In the case of IR, the structure of words embedded in the reports (and constitutive of the category of IT) contributes by changing the reports' semantic orientation over time by providing ongoing flows of meaning. This mutual relationship between IT and integrated reports emerges from our analysis and is witnessed by the fact that changes in the composition of capitals and their linkages during the years correspond to changing semantic orientations of the reports within a pattern characterized by the shift from a financial and managerial perspective (internally-oriented) to a commercial and relational perspective (externally-oriented).

The presence of several different linkages between multiple (and changing) capitals also contributes to explain the vagueness of the concept of IT as highlighted by previous research (Dumay *et al.*, 2017). Our findings suggest that, while it would be impossible to reduce such vagueness (due to the dynamic meaning of IT and the impossibility to achieve its definite conceptualization), the study of the vocabulary structure of the reports can make the meaning of IT visible at least temporarily, to engage organizations' managers and stakeholders in sense-making. IR is a signifying practice that provides different images of an organization's value production process, which are drawn by representing the different capitals (on which this process is focused) and the paths through which these capitals are connected. In this way, IR provides managers and stakeholders with a larger number of vocabularies which give them options for constructing the meaning of actions they carry out giving substance to the concept of IT.

Therefore, IR is a practice more likely able "to give opportunities to reveal the latent opportunities in what might otherwise seem like blatant threats" because of the difficulties of building a common understanding between subjects (Weick, 1995, p. 184). While studies have shown how the concept of IR can be interpreted in different ways because of the different perceptions of IR preparers (Feng *et al.*, 2017; Lai *et al.*, 2017), our analysis supports the idea that IR is a signifying practice that, through an ensemble of vocabularies, produces at different times a flow of meanings able to build (at least temporarily) a common understanding of vague concepts such as IT.

6. Conclusions

This paper has explored the main objects represented in integrated reports, which contribute to make visible and material the concept of IT. Most literature and practitioners have looked at IT mainly within a managerial perspective, able to facilitate or hamper the drawing of integrated reports and their quality (Dumay *et al.*, 2017; Feng *et al.*, 2017; International Integrated Reporting Council, 2016). Few scholars and practitioners have analyzed IR as a signifying practice, which contributes to construct and communicate the concept of IT to reach a common understanding of it (Busco *et al.*, 2017, 2018).

By drawing on a vocabulary approach, we have shown that integrated reports represent IT as a dynamic concept in its meaning due to the possible change and variable weight of each element that constitutes the different objects represented and their connections (i.e. capitals and their linkages in our case). It is through these objects and their connections that the concept of IT is shared and takes shape. Both the different systems of words constitutive of the various capitals and their connections, which are represented in integrated reports, make these objects follow different paths. This, in turn, makes the meaning of IT in constant change, therefore, without ever reaching a definite conceptualization. This highlighted dynamism of the concept of IT makes the analysis of integrated reports an important way to understand this concept, its evolution over time and, consequently, for reaching a common understanding. However, the proposed analysis that has been demonstrated in this paper

could be useful to evaluate a possible decoupling between what is represented and what is really thought about within an organization.

Therefore, this paper contributes a further development in understanding to the literature on IR practices by responding to the call for more research into the specific concept of IT (Dumay and Dai, 2017; Feng *et al.*, 2017; Oliver *et al.*, 2016; Rinaldi *et al.*, 2018). While previous studies have analyzed the content of integrated reports for investigating various features and levels of compliance to the framework (Lai *et al.*, 2017), we have focused, in particular, on the meaning of IT, which manifests itself within and through the reports. In so doing, we also contribute to add insight into the research on IR as signifying practices (Busco *et al.*, 2017, 2018; Hall, 1997) by showing that IR gives meaning to the concept of IT while representing the objects around which IT is commonly understood within organizations. Furthermore, we reinforce the findings about IT as a “vague” concept (Dumay *et al.*, 2017) and the need for more insights into what IT means in practice (Oliver *et al.*, 2016). However, in so doing, we still emphasize the dynamic perspective through which IT should be analyzed according to the evolutionary paths of interactions among the objects to which it may refer. Specifically, we show that, rather than providing a clear definition of IT, we need a better comprehension of how its unfolding (and never complete/definite) meaning is constructed through the words included in the reports.

In this vein, this paper also contributes to paving the way for a linguistic approach to analyzing integrated reports. While other studies have shown the complex nature of IT, originating from internal processes and cultures (Dumay and Dai, 2017; McNally *et al.*, 2017), our findings suggest the possibility of investigating IT by looking at the systems of words and their meanings which make IT concrete and meaningful within the reports. This aspect is particularly relevant due to its related practical implications with the different subjects directly and/or indirectly involved in and/or interested in integrated reports. Preparers of integrated reports can give more attention to the lemmas used in drafting integrated reports and, especially, their narrative sections as they contain meanings and linkages among the dimensions characterizing IT, which are mobilized through integrated reports. Indeed, the study of vocabularies may help preparers in understanding whether and how integrated reports represent their organization in a holistic way and if the perception coming from this representation is coherent with its different stakeholders’ needs and the environment in which the organization operates, therefore, allowing it to reach an increasing level of legitimation. Users, and especially readers, of integrated reports, have at their disposal a technique able to provide a more analytical understanding of the content of these documents. The use of the technique proposed in this paper permits a reduction in the complexity of integrated reports by singling out the different components in the concept under investigation, in our case IT, and the linkages among them. In addition, the use of the technique gives users of integrated reports the possibility to make more reliable comparisons among those published in different time periods or by different subjects. Finally, IIRC could analyze integrated reports published by different organizations to evaluate their quality, as well as the effectiveness of the IIRC’s guidelines. For example, the vocabulary approach to integrated reports could be useful for IIRC to understand the existing level of connectivity within the reports and, consequently, develop new and/or updated guidelines. In so doing, we show the potential for a vocabulary approach to IR and IT from a practice perspective as well.

Although our findings derive from a limited number of selected integrated reports, the research design we followed could be extended to a larger number of reports. In so doing, by using a vocabulary approach, further studies could investigate whether other kinds of capital emerge from the analysis, as well as the possible reasons behind different

evolutionary paths of the meanings and materialization of IT (e.g. the organizations' sector, country, compulsory or voluntary adoption of IR, etc.). Moreover, future studies could extend the analysis of the vocabulary structure of integrated reports by focusing on the combination of word-to-word relationships with word-to-example relationships. Finally, further research could combine the analysis of the vocabulary structures of the reports with field studies to explore, on the one hand, how words make IT concrete in specific organizational contexts and, on the other hand, how the production process of the reports (e.g. internally produced or outsourced to a consultant) could affect the materialization of IT.

Notes

1. The chi-square test is calculated by the software T-Lab for verifying the significance of a specific lemma (e.g. "x") in a specific cluster (e.g. "y"). Chi square test is run on the 2x2 tables constructed considering the occurrences of the lemma "x" in the cluster "y", the occurrences of the lemma "x" in clusters other than "y", the occurrences of lemmas other than "x" in the cluster "y" and the occurrences of lemmas other than "x" in clusters other than "y". The threshold value is 3.84 ($df = 1; p = 0.05$).
2. Using the software T-Lab, we selected the option "few links" to obtain a simplified and understandable visualization of the links between different themes. T-Lab discarded edges with a probability value lower than 0.175.

References

- Adams, C.A. (2004), "The ethical, social and environmental reporting-performance portrayal gap", *Accounting, Auditing and Accountability Journal*, Vol. 17 No. 5, pp. 731-757.
- Al-Htaybat, K. and von Alberti-Alhtaybat, L. (2018), "Integrated thinking leading to integrated reporting: case study insights from a global player", *Accounting, Auditing and Accountability Journal*, Vol. 31 No. 5, pp. 1435-1460.
- Alvesson, M. and Kärreman, D. (2000), "Taking the linguistic turn in organizational research: challenges, responses, consequences", *The Journal of Applied Behavioral Science*, Vol. 36 No. 2, pp. 136-158.
- Barnabè, F., Giorgino, M.C. and Kunc, M. (2019), "Visualizing and managing value creation through integrated reporting practices: a dynamic resource-based perspective", *Journal of Management and Governance*, Vol. 23 No. 2, pp. 537-575.
- Beattie, V. (2014), "Accounting narratives and the narrative turn in accounting research: issues, theory, methodology, methods and a research framework", *The British Accounting Review*, Vol. 46 No. 2, pp. 111-134.
- Black Sun (2012), "Building the business case for integrated reporting", IIRC (The International Integrated Reporting Council), available at: <https://integratedreporting.org/wp-content/uploads/2012/11/BUILDING-THE-BUSINESS-CASE-FOR-INTEGRATED-REPORTING.pdf> (accessed 5 April 2020).
- Brennan, N.M., Guillamon Saorin, E. and Pierce, A. (2009), "Methodological insights: impression management: developing and illustrating a scheme of analysis for narrative disclosures – a methodological note", *Accounting, Auditing and Accountability Journal*, Vol. 22 No. 5, pp. 789-832.
- Brown, J. and Dillard, J. (2014), "Integrated reporting: on the need for broadening out and opening up", *Accounting, Auditing and Accountability Journal*, Vol. 27 No. 7, pp. 1120-1156.
- Burritt, R.L. and Schaltegger, S. (2010), "Sustainability accounting and reporting: fad or trend?", *Accounting, Auditing and Accountability Journal*, Vol. 23 No. 7, pp. 829-846.

- Busco, C. Granà, F. and Quattrone, P. (2017), "Integrated thinking. Aligning purpose and the business model to market opportunities and sustainable performance", CIMA Research Executive Summary, available at: www.cimaglobal.com/Documents/ResearchandInsight/IntegratedThinkingReportvol13issue3.pdf (accessed 8 January 2020).
- Busco, C., Frigo, M.L., Riccaboni, A. and Quattrone, P. (Eds) (2013), "Integrated reporting", *Concepts and Cases That Redefine Corporate Accountability*, Springer, London.
- Busco, C., Giovannoni, E., Granà, F. and Izzo, M.F. (2018), "Making sustainability meaningful: aspirations, discourses and reporting practices", *Accounting, Auditing and Accountability Journal*, Vol. 31 No. 8, pp. 2218-2246.
- Busco, C., Malafronte, I., Pereira, J. and Starita, M.G. (2019), "The determinants of companies' levels of integration: does one size fit all?", *The British Accounting Review*, Vol. 51 No. 3, pp. 277-298.
- Caglio, A., Melloni, G. and Perego, P. (2019), "Informational content and assurance of textual disclosures: evidence on integrated reporting", *European Accounting Review*, Vol. 29 No. 1, pp. 55-83, doi: [10.1080/09638180.2019.1677486](https://doi.org/10.1080/09638180.2019.1677486).
- Carley, K. (1993), "Coding choices for textual analysis: a comparison of content analysis and map analysis", *Sociological Methodology*, Vol. 23, pp. 75-126.
- Carley, K. (1994), "Extracting culture through textual analysis", *Poetics*, Vol. 22 No. 4, pp. 291-312.
- Carley, K.M. and Kaufer, D.S. (1993), "Semantic connectivity: an approach for analyzing symbols in semantic networks", *Communication Theory*, Vol. 3 No. 3, pp. 183-213.
- Chartered Institute of Management Accountants (2017), "Integrated thinking. Aligning purpose and the business model to market opportunities and sustainable performance", CIMA Research Executive Summary, available at: www.cimaglobal.com/Documents/ResearchandInsight/IntegratedThinkingReportvol13issue3.pdf (accessed 8 January 2020).
- Checkland, P. (1981), *Systems Thinking. Systems Practice*, John Wiley and Sons, Chichester.
- Child, J. and Rodrigues, S.B. (2011), "How organizations engage with external complexity: a political action perspective", *Organization Studies*, Vol. 32 No. 6, pp. 803-824.
- Churet, C. and Eccles, R.G. (2014), "Integrated reporting, quality of management, and financial performance", *Journal of Applied Corporate Finance*, Vol. 26 No. 1, pp. 8-16.
- Contrafatto, M. and Burns, J. (2013), "Social and environmental accounting, organisational change and management accounting: a processual view", *Management Accounting Research*, Vol. 24 No. 4, pp. 349-365.
- de Villiers, C., Venter, E.R. and Hsiao, P.C.K. (2017), "Integrated reporting: background, measurement issues, approaches and an agenda for future research", *Accounting and Finance*, Vol. 57 No. 4, pp. 937-959.
- Deloitte (2015), "Integrated reporting as a driver for integrated thinking?", available at: www2.deloitte.com/content/dam/Deloitte/nl/Documents/risk/deloitte-nl-risk-integrated-reporting-a-driver-for-integrated-thinking.pdf (accessed 8 January 2020).
- Diesner, J. and Carley, K.M. (2005), "Revealing social structure from texts: meta-matrix text analysis as a novel method for network text analysis", in Narayanan, V.K. and Armstrong, D.J. (Eds), *Causal Mapping for Research in Information Technology*, Idea Group Publishing, Hershey, pp. 81-108.
- DiMaggio, P.J. and Powell, W.W. (1983), "The iron cage revisited: institutional isomorphism and collective rationality in organizational fields", *American Sociological Review*, Vol. 48 No. 2, pp. 147-160.
- Dossi, A. and Patelli, L. (2010), "You learn from what you measure: financial and non-financial performance measures in multinational companies", *Long Range Planning*, Vol. 43 No. 4, pp. 498-526.
- Du Toit, E. (2017), "The readability of integrated reports", *Meditari Accountancy Research*, Vol. 25 No. 4, pp. 629-653.

- Dumay, J. and Dai, T. (2017), "Integrated thinking as a cultural control?", *Meditari Accountancy Research*, Vol. 25 No. 4, pp. 574-604.
- Dumay, J., Bernardi, C., Guthrie, J. and La Torre, M. (2017), "Barriers to implementing the international integrated reporting framework a contemporary academic perspective", *Meditari Accountancy Research*, Vol. 25 No. 4, pp. 461-480.
- Feng, T., Cummings, L. and Tweedie, D. (2017), "Exploring integrated thinking in integrated reporting – an exploratory study in Australia", *Journal of Intellectual Capital*, Vol. 18 No. 2, pp. 330-353.
- Gabriel, K.R. (1971), "The biplot graphic display of matrices with application to principal component analysis", *Biometrika*, Vol. 58 No. 3, pp. 453-467.
- Gibassier, D., Rodrigue, M. and Arjaliès, D.L. (2018), "Integrated reporting is like god: no one has met him, but everybody talks about him': the power of myths in the adoption of management innovations", *Accounting, Auditing and Accountability Journal*, Vol. 31 No. 5, pp. 1349-1380.
- Gray, R. (2002), "The social accounting project and accounting organizations and society. Privileging engagement, imaginings, new accountings and pragmatism over critique?", *Accounting, Organizations and Society*, Vol. 27 No. 7, pp. 687-708.
- Greenacre, M.J. (1984), *Theory and Applications of Correspondence Analysis*, Academic Press, New York, NY.
- Guthrie, J. and Dumay, J. (2015), "New frontiers in the use of intellectual capital in the public sector", *Journal of Intellectual Capital*, Vol. 16 No. 2, pp. 258-266.
- Guthrie, J., Manes-Rossi, F. and Orelli, R.L. (2017), "Integrated reporting and integrated thinking in Italian public sector organizations", *Meditari Accountancy Research*, Vol. 25 No. 4, pp. 553-573.
- Hall, S. (1997), "The spectacle of the 'other'", in Hall, S. (Ed.), *Representation: Cultural Representations and Signifying Practices*, SAGE, Thousand Oaks, CA, London and New Delhi, pp. 225-239.
- Hamel, G. (2006), "The why, what, and how of management innovation", *Harvard Business Review*, Vol. 84 No. 2, pp. 72-84.
- Higgins, C., Stubbs, W. and Love, T. (2014), "Walking the talk(s): organisational narratives of integrated reporting", *Accounting, Auditing and Accountability Journal*, Vol. 27 No. 7, pp. 1090-1119.
- Hoque, Z. (2014), "20 years of studies on the balanced scorecard: trends, accomplishments, gaps and opportunities for future research", *The British Accounting Review*, Vol. 46 No. 1, pp. 33-59.
- International Integrated Reporting Council (2013), "The international <IR> framework", available at: <https://integratedreporting.org/wp-content/uploads/2015/03/13-12-08-THE-INTERNATIONAL-IR-FRAMEWORK-2-1.pdf> (accessed 8 January 2020).
- International Integrated Reporting Council (2016), "Creating value. The cyclical power of integrated thinking and reporting", available at: https://integratedreporting.org/wp-content/uploads/2017/05/CreatingValue_IntegratedThinkingK1.pdf (accessed 8 January 2020).
- Jones, C. and Livne-Tarandach, R. (2008), "Designing a frame: rhetorical strategies of architects", *Journal of Organizational Behavior*, Vol. 29 No. 8, pp. 1075-1099.
- Kaplan, R.S. and Norton, D.P. (1992), "The balanced scorecard - measures that drive performance", *Harvard Business Review*, pp. 71-79.
- Lai, A., Melloni, G. and Stacchezzini, R. (2017), "What does materiality mean to integrated reporting preparers? An empirical exploration", *Meditari Accountancy Research*, Vol. 25 No. 4, pp. 533-552.
- Lai, A., Melloni, G. and Stacchezzini, R. (2018), "Integrated reporting and narrative accountability: the role of preparers", *Accounting, Auditing and Accountability Journal*, Vol. 31 No. 5, pp. 1381-1405.
- Lancia, F. (2019), "T-lab plus user's manual. Tool for text analysis", available at: www.tlab.it
- Lebart, L. and Salem, A. (1988), *Statistique Des Données Textuelles: questions Ouvertes et Lexicométrie*, Dunod, Paris.

- Lebart, L., Morineau, A. and Warwick, K.M. (1984), *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices*, John Wiley and Sons, New York, NY.
- Lebart, L., Salem, A. and Berry, L. (1998), "Exploring textual data", *Springer Science and Business Media*, Vol. 4, Springer.
- Lengnick-Hall, C.A. and Sanders, M.M. (1997), "Designing effective learning systems for management education: student roles, requisite variety, and practicing what we teach", *Academy of Management Journal*, Vol. 40 No. 6, pp. 1334-1368.
- Loewenstein, J. and Ocasio, W.C. (2005), "Vocabularies of organizing: how language links culture, cognition, and action in organizations", McCombs Working Paper No. OSSM-03-05, available at SSRN: <https://ssrn.com/abstract=813344> or doi: 10.2139/ssrn.813344 doi: [10.2139/ssrn.813344](https://doi.org/10.2139/ssrn.813344).
- Loewenstein, J., Ocasio, W.C. and Jones, C. (2012), "Vocabularies and vocabulary structure: a new approach linking categories, practices, and institutions", *Academy of Management Annals*, Vol. 6 No. 1, pp. 41-86.
- Loughran, T. and McDonald, B. (2016), "Textual analysis in accounting and finance: a survey", *Journal of Accounting Research*, Vol. 54 No. 4, pp. 1187-1230.
- McEwan, J.A. and Schlich, P. (1991), "Correspondence analysis in sensory evaluation", *Food Quality and Preference*, Vol. 3 No. 1, pp. 23-36.
- McNally, M., Cerbone, D. and Maroun, W. (2017), "Exploring the challenges of preparing an integrated report", *Meditari Accountancy Research*, Vol. 25 No. 4, pp. 481-504.
- Melloni, G., Caglio, A. and Perego, P. (2017), "Saying more with less? Disclosure conciseness, completeness and balance in integrated reports", *Journal of Accounting and Public Policy*, Vol. 36 No. 3, pp. 220-238.
- Milne, M.J., Tregidga, H. and Walton, S. (2009), "Words not actions! The ideological role of sustainable development reporting", *Accounting, Auditing and Accountability Journal*, Vol. 22 No. 8, pp. 1211-1257.
- Modell, S. (2009), "Institutional research on performance measurement and management in the public sector accounting literature: a review and assessment", *Financial Accountability and Management*, Vol. 25 No. 3, pp. 277-303.
- Monfardini, P., Barretta, A.D. and Ruggiero, P. (2013), "Seeking legitimacy: social reporting in the healthcare sector", *Accounting Forum*, Vol. 37 No. 1, pp. 54-66.
- Montecalvo, M., Farneti, F. and de Villiers, C. (2018), "The potential of integrated reporting to enhance sustainability reporting in the public sector", *Public Money and Management*, Vol. 38 No. 5, pp. 365-374.
- Mouritsen, J. (1998), "Driving growth: economic value added versus intellectual capital", *Management Accounting Research*, Vol. 9 No. 4, pp. 461-482.
- O'Donovan, G. (2002), "Environmental disclosures in the annual report: extending the applicability and predictive power of legitimacy theory", *Accounting, Auditing and Accountability Journal*, Vol. 15 No. 3, pp. 344-371.
- Oliver, J., Vesty, G. and Brooks, A. (2016), "Conceptualising integrated thinking in practice", *Managerial Auditing Journal*, Vol. 31 No. 2, pp. 228-248.
- Poulis, K. and Poulis, E. (2016), "Problematising fit and survival: transforming the law of requisite variety through complexity misalignment", *Academy of Management Review*, Vol. 41 No. 3, pp. 503-527.
- Rinaldi, L., Unerman, J. and de Villiers, C. (2018), "Evaluating the integrated reporting journey: insights, gaps and agendas for future research", *Accounting, Auditing and Accountability Journal*, Vol. 31 No. 5, pp. 1294-1318.
- Salton, G. (1989), *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Vol. 169, Addison-Wesley, Reading.

-
- Savaresi, S.M. and Boley, D.L. (2001), "On the performance of bisecting K-means and PDDP", *Proceedings of the 2001 SIAM International Conference on Data Mining, Society for Industrial and Applied Mathematics*, pp. 1-14.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003), "Cytoscape: a software environment for integrated models of biomolecular interaction networks", *Genome Research*, Vol. 13 No. 11, pp. 2498-2504.
- Steinbach, M., Karypis, G. and Kumar, V. (2000), "A comparison of document clustering techniques", *KDD Workshop on Text Mining*, Vol. 400 No. 1.
- Stent, W. and Dowler, T. (2015), "Early assessments of the gap between integrated reporting and current corporate reporting", *Meditari Accountancy Research*, Vol. 23 No. 1, pp. 92-117.
- Trobia, A. and Lo Verde, F.M. (2017), "Italian amateur pop-rock musicians on Facebook: mixed methods and new findings in music-making research", in Mantie, R. and Smith, G.D. (Eds), *The Oxford Handbook of Music Making and Leisure*, Oxford University Press, Oxford, pp. 491-518.
- Vasconcelos, F.C. and Ramirez, R. (2011), "Complexity in business environments", *Journal of Business Research*, Vol. 64 No. 3, pp. 236-241.
- Weick, K.E. (1995), *Sensemaking in Organizations*, Sage, Thousand Oaks.
- Wild, S. and Van Staden, C. (2013), "Integrated reporting: initial analysis of early reporters – an institutional theory approach", *7th Asia Pacific Interdisciplinary Research in Accounting Conference, Kobe*.
- World Intellectual Capital Initiative (2013), "Connectivity. Background paper for <IR>", International Integrated Reporting Council, available at: <https://integratedreporting.org/wp-content/uploads/2012/11/BUILDING-THE-BUSINESS-CASE-FOR-INTEGRATED-REPORTING.pdf> (accessed 5 April 2020).
- Zappettini, F. and Unerman, J. (2016), "'Mixing' and 'bending': the recontextualisation of discourses of sustainability in integrated reporting", *Discourse and Communication*, Vol. 10 No. 5, pp. 521-542.

Corresponding author

Pasquale Ruggiero can be contacted at: pasquale.ruggiero@unisi.it

Organizations' names	Country/Industry	Reports: Labels, years and no. of pages
Bankmecu (Bank Australia since 2015)	– Australia – Industry: banking	– Annual Report 2012 (13 pp.) – Corporate Report 2013 (13 pp.) – Corporate Report 2014 (6 pp.) – Corporate Report 2015 (62 pp.) – Corporate Report 2016 (57 pp.) – Corporate Report 2017 (68 pp.) – Corporate Report 2018 (81 pp.)
Liberty Holdings	– South Africa – Industry: insurance	– Integrated Annual Report 2012 (381 pp.) – Integrated Annual Report 2013 (154 pp.) – Integrated Report 2014 (105 pp.) – Integrated Report 2015 (85 pp.) – Integrated Report 2016 (85 pp.) – Integrated Report 2017 (82 pp.) – Integrated Report 2018 (87 pp.)
Standard Bank Group	– South Africa – Industry: banking and investment services	– Annual Integrated Report 2012 (178 pp.) – Annual Integrated Report 2013 (188 pp.) – Annual Integrated Report 2014 (196 pp.) – Annual Integrated Report 2015 (129 pp.) – Annual Integrated Report 2016 (129 pp.) – Annual Integrated Report 2017 (125 pp.) – Annual Integrated Report 2018 (105 pp.)
Stockland*	– Australia – Industry: real estate	– Annual Review 2012 (56 pp.) – Annual Review 2013 (57 pp.) – Annual Review 2015 (64 pp.) – Annual Review 2016 (76 pp.) – Annual Review 2017 (67 pp.) – Annual Review 2018 (64 pp.)
Truworths	– South Africa – Industry: Apparel and accessories retailer	– Integrated Annual Report 2012 (138 pp.) – Integrated Annual Report 2013 (134 pp.) – Integrated Report 2014 (138 pp.) – Integrated Report 2015 (110 pp.) – Integrated Report 2016 (109 pp.) – Integrated Report 2017 (68 pp.) – Integrated Report 2018 (69 pp.)

Notes: *Stockland's Annual Review 2014 has not been included in the analysis because the report was not available online. [Table A1](#) shows the empirical material which includes 34 integrated reports published by five organizations from 2012 to 2018. The reports have been selected from the IIRC online database, section "recognized reports." Reports have been selected if they were awarded in 2013 and 2014 and then have been continuously published in the following years (independently from industry or type of organization). See Section 3 of the paper

Table A1.
The reports analyzed

Source: Our empirical material

LEMMA	CONT-1	CONT-2	CONT-3	CONT-4	CONT-5	CONT-6
Client	0.025	0.0039	0.0027	0	0.0014	0.0036
Share	0.0092	0.0004	0.0015	0.0006	0.0005	0.0002
Company	0.0088	0.0009	0.0014	0.0001	0.0005	0.0009
Limited	0.0083	0.0002	0.002	0.0005	0.0008	0.0006
Group	0.0076	0.0033	0.0008	0.0032	0.0007	0.0001
Liability	0.0075	0.0107	0.0005	0	0.0001	0
Holding	0.0073	0	0.0008	0.0004	0.0001	0
Policyholder	0.0059	0.0097	0.0005	0.0002	0.0003	0
Australia	0.0058	0.0008	0.0002	0.008	0	0.0003
Digital	0.0054	0.0002	0.0009	0.0004	0.0003	0.0001
Shareholder	0.0053	0.0005	0.001	0	0.0038	0
Insurance	0.0046	0.0039	0.0001	0	0.0004	0
Customer	0.0045	0.0018	0.0038	0.0006	0	0.0091
Fair_value	0.0042	0.0053	0	0	0.0004	0.0002
E-commerce	0.0041	0.0011	0.0012	0.0014	0.0003	0.0001
Community	0.004	0.0001	0.0017	0	0.0044	0.0003
Contract	0.004	0.008	0.0012	0.0001	0.0024	0
Planet	0.0039	0.002	0.0075	0.0112	0.0006	0.0012
Outcome	0.0038	0.0007	0.0001	0.0001	0	0.0011
People	0.0038	0	0.0002	0.0022	0	0.0001
Risk	0.0038	0.0033	0.0022	0.0001	0.0017	0
Create	0.0036	0.0005	0.0003	0.0003	0.0009	0.0002
Region	0.0036	0.0005	0	0.0041	0	0.0002
Taxation	0.0036	0.007	0.0011	0.0001	0.0017	0.0002
Liability	0.0075	0.0107	0.0005	0	0.0001	0
Property	0.0025	0.01	0.0035	0.0003	0.0004	0.0001
Integrated_report	0.0002	0.0098	0.0078	0.01	0.0155	0.0046
Policyholder	0.0059	0.0097	0.0005	0.0002	0.0003	0
Fund	0.0018	0.0081	0.0003	0.0001	0.0014	0.0001
Contract	0.004	0.008	0.0012	0.0001	0.0024	0
Appendix	0.0025	0.0075	0.0124	0.0005	0.0022	0.0009
Chairman	0.001	0.007	0.0003	0.0001	0.0004	0.0003
Taxation	0.0036	0.007	0.0011	0.0001	0.0017	0.0002
Claim	0.0008	0.0062	0.0005	0	0.0007	0
Per_cent	0	0.0059	0.0091	0.0006	0.0079	0.0003
Investment	0.0018	0.0056	0.0027	0.0002	0.0016	0.0001
XXX (ceo name)	0.0011	0.0054	0.0029	0.0008	0.0038	0.0002
Fair_value	0.0042	0.0053	0	0	0.0004	0.0002
Explain	0.0001	0.0052	0.0013	0.0001	0.0099	0
Ce	0	0.0051	0.0006	0.0005	0	0.003
Director	0.0024	0.0051	0.0011	0	0	0
Option	0.0008	0.0051	0	0.0003	0	0
Proprietary	0.0023	0.0051	0.0002	0	0.0007	0
Derivative	0.0025	0.0049	0.0003	0	0.0003	0.0002
Reinsurance	0.0022	0.0044	0.0004	0	0.0002	0
Annuity	0.0017	0.0043	0.0002	0	0.0006	0
Lock-in	0.0002	0.0043	0.0017	0.0007	0.0083	0.0001
Proxy	0.0018	0.0042	0.0025	0.0002	0	0

Notes: Table A2 reports a part of the table with the lemmas' contribution to inertia value, i.e. the percentage of information that each lemma explains about each of the six dimensions representing the latent meaning of the corpus (from cont-1 to cont-6). Lemmas are sorted by the contribution to inertia value for the first dimension (cont-1, i.e. the X-axis of the biplot shown in Section 4.1) and for the second dimension (cont-2, i.e. the Y-axis of the biplot shown in Section 4.1). The lemmas with the highest contribution to inertia values help in the interpretation of the dimensions' latent meanings

Source: Part of the total output

Table A2.
The lemmas' contributions to factorial axes inertia for the two dimensions

<i>Lemmas FC1</i>	<i>Occurrences in FC1</i>	<i>Total occurrences</i>	χ^2	<i>p-value</i>
Asset	2,389	3,346	3,184.55	0.000
Increase	2,132	3,018	2,772.36	0.000
Million	1,996	2,748	2,763.44	0.000
Income	1,317	1,490	2,731.44	0.000
Rm	1,177	1,348	2,383.77	0.000
Cost	1,616	2,162	2,381.70	0.000
Rate	1,266	1,571	2,186.14	0.000
Liability	1,082	1,302	1,987.97	0.000
Earning	1,161	1,450	1,978.12	0.000
Billion	1,122	1,395	1,929.33	0.000
Growth	1,986	3,205	1,872.90	0.000
Interest	1,437	2,068	1,796.19	0.000
Profit	1,018	1,256	1,777.96	0.000
Credit	1,445	2,180	1,613.96	0.000
Cash	1,232	1,741	1,605.44	0.000
Loss	834	1,035	1,438.85	0.000
Revenue	737	902	1,307.86	0.000
Margin	676	793	1,305.51	0.000
Total	1,484	2,460	1,301.45	0.000
Impairment	548	586	1,262.08	0.000
Fund	1,281	2,050	1,232.83	0.000
Trading	794	1,058	1,178.69	0.000
Loan	734	949	1,162.39	0.000
Return	999	1,509	1,111.39	0.000
Debt	603	733	1,084.65	0.000
<i>Lemmas HC2</i>	<i>Occurrences in HC2</i>	<i>Total occurrences</i>	χ^2	<i>p-value</i>
Award	1,587	1,896	5,143.38	0.000
Employee	2,251	3,741	4,114.25	0.000
Incentive	1,047	1,183	3,700.00	0.000
Performance	2,360	4,772	2,872.16	0.000
Scheme	1,021	1,355	2,784.96	0.000
Remuneration	1,584	2,854	2,462.80	0.000
Reward	613	742	1,942.48	0.000
Vesting	466	510	1,727.99	0.000
Right	687	965	1,700.12	0.000
Defer	574	774	1,522.98	0.000
Senior	558	823	1,267.14	0.000
Plan	984	2,132	1,015.65	0.000
Stakeholder	808	1,661	942.96	0.000
Pay	389	553	942.26	0.000
Bonus	275	333	870.38	0.000
Value	1,674	4,715	829.75	0.000
Individual	531	955	826.45	0.000
Variable	276	369	742.52	0.000
Long-term	684	1,461	729.82	0.000
Engagement	612	1,256	716.53	0.000
(name of a manager)	203	238	676.24	0.000
Grant	292	430	664.58	0.000
Executive	949	2,389	661.98	0.000
PRP	164	177	620.90	0.000
Retention	231	313	607.50	0.000
<i>Lemmas SRC3</i>	<i>Occurrences in SRC3</i>	<i>Total occurrences</i>	χ^2	<i>p-value</i>
community	1,297	1,861	7,659.99	0.000
customer	1,658	5,150	2,992.70	0.000
Project	631	1,194	2,539.42	0.000

Table A3.
The first 25 lemmas
with the highest chi-
square for
each theme

(continued)

residential	331	482	1,913.21	0.000
living	340	532	1,784.47	0.000
Centre	461	923	1,707.19	0.000
Australia	402	741	1,679.10	0.000
Resident	221	263	1,658.05	0.000
Program	269	375	1,646.80	0.000
Energy	340	565	1,643.19	0.000
Green	196	260	1,279.96	0.000
Development	749	2,454	1,217.11	0.000
Village	157	189	1,160.77	0.000
Retirement	371	834	1,152.11	0.000
Star	170	222	1,132.86	0.000
Education	222	375	1,048.52	0.000
Water	191	305	974.74	0.000
Satisfaction	215	376	967.28	0.000
Wellbeing	142	183	962.47	0.000
Biodiversity	129	167	869.26	0.000
Liveability	101	111	836.82	0.000
Housing	147	217	834.06	0.000
Renewable	128	183	757.77	0.000
Home	178	333	727.38	0.000
Solar	93	126	590.78	0.000
<i>Lemmas PC4</i>	<i>Occurrences in PC4</i>	<i>Total occurrences</i>	χ^2	<i>p-value</i>
Report	3,116	6,022	6,884.36	0.000
Risk	2,893	6,723	4,523.68	0.000
Governance	1,237	2,346	2,810.42	0.000
Assurance	558	705	2,403.81	0.000
Framework	668	985	2,286.99	0.000
Compliance	592	1,009	1,609.48	0.000
Information	856	1,984	1,336.55	0.000
Monitor	487	868	1,228.53	0.000
Internal	441	829	1,015.35	0.000
Risk_management	332	537	982.23	0.000
Group	3,216	13,601	966.08	0.000
Material	529	1,137	955.46	0.000
Audit	522	1,134	923.86	0.000
Control	452	958	838.01	0.000
Review	919	2,733	824.99	0.000
Disclosure	276	449	808.92	0.000
Regulatory	490	1,109	801.04	0.000
Detail	464	1,064	739.12	0.000
Requirement	563	1,419	738.64	0.000
Code	336	657	721.55	0.000
Defence	136	153	689.84	0.000
Standard	675	1,917	677.15	0.000
Refer	348	734	650.89	0.000
Sustainability	451	1,106	626.68	0.000
Principle	347	750	619.78	0.000
<i>Lemmas OC5</i>	<i>Occurrences in OC5</i>	<i>Total occurrences</i>	χ^2	<i>p-value</i>
Director	3,696	4,272	16,449.23	0.000
Board	3,055	4,162	10,646.49	0.000
Committee	2,728	3,777	9,267.74	0.000
Non-executive	1,176	1,240	5,934.88	0.000
Company	1,682	2,427	5,346.27	0.000
Chairman	1,190	1,440	4,935.18	0.000
Appoint	948	1,006	4,740.23	0.000
Meeting	1,029	1,219	4,400.42	0.000

(continued)

Table A3.

MEDAR
29,4

774

Independent	776	1,105	2,513.73	0.000
Member	656	949	2,072.29	0.000
Executive	1,171	2,389	2,038.56	0.000
Chief_executive	521	656	2,035.39	0.000
Limited	1,170	2,397	2,020.37	0.000
Chief	560	828	1,708.39	0.000
Ltd	475	671	1,557.13	0.000
Secretary	312	333	1,546.10	0.000
General	501	757	1,475.92	0.000
Proxy	306	341	1,426.91	0.000
Officer	544	926	1,314.21	0.000
Attendance	218	225	1,131.70	0.000
Vote	330	458	1,111.66	0.000
Holding	720	1,567	1,101.71	0.000
Pty	227	247	1,094.70	0.000
Affair	223	242	1,079.45	0.000
Shareholder	996	2,667	964.84	0.000
<i>Lemmas CC6</i>		<i>Total occurrences</i>	χ^2	<i>p-value</i>
Store	1,482	1,852	7,062.24	0.000
Client	1,480	2,345	4,979.54	0.000
Brand	948	1,198	4,442.91	0.000
Fashion	710	930	3,161.19	0.000
Product	1,056	2,122	2,408.12	0.000
Emporium	416	453	2,390.11	0.000
Offering	422	532	1,982.63	0.000
Digital	394	487	1,902.94	0.000
Platform	388	488	1,828.76	0.000
Merchandise	483	787	1,552.62	0.000
Solution	362	554	1,281.19	0.000
Service	881	2,391	1,126.38	0.000
Offer	388	676	1,121.31	0.000
Africa	882	2,447	1,075.10	0.000
Customer	1,486	5,150	1,055.35	0.000
E-commerce	191	221	1,010.36	0.000
Capability	398	770	967.04	0.000
System	515	1,158	957.68	0.000
Identity	201	257	925.31	0.000
Leverage	265	410	922.25	0.000
Launch	322	580	881.41	0.000
Channel	244	374	861.29	0.000
Range	413	902	810.22	0.000
Man	168	209	803.78	0.000
Kid	127	135	753.28	0.000

Notes: Table A3 shows the first 25 lemmas with the highest chi-square for each theme (FC1-financial capital, HC2-human capital, SRC3-social-relational capital, PC4-process capital, OC5-organizational capital, CC6-commercial capital) identified through the thematic analysis of elementary contexts. The chi-square test is calculated by the software T-Lab for verifying the significance of a specific lemma in a specific cluster. For each lemma, the table also indicates the number of occurrences in the theme and the total number of occurrences in the reports. As shown in the table, the analysis is statistically significant at 0.05 level (p -value ≤ 0.05). See Section 4.2 in the paper

Table A3.

Source: Part of the total output