

One More Tweet: Firms Challenge a Sustainable Future

MARCO TREGUA

marco.tregua@unina.it

University of Naples Federico II

ANNA D'AURIA

anna.dauria@unina.it

University of Naples Federico II

Abstract

In the last decades, researchers have been provided with a huge amount of data thanks to the diffusion of online sources. Additionally, more companies are issuing reports and documents to share information with stakeholders about their sustainable approach to both strengthen and encourage people to adopt a similar approach. To support researchers in managing the increasing quantity of information, several tools have been provided for text mining studies, such as sentiment analysis, semantic analysis, and content analysis. We proposed to analyse the usefulness of automated and semi-automated techniques on a dataset composed of more than 875,000 tweets posted by the companies that Forbes (2020) considers to be the most sustainable. We chose to focus on sustainability because it is a topic of interest to the global community, as revealed by the significant amount of attention that companies are paying to it. In detail, we performed a double-step analysis: firstly, a comparison between *exact words* and *stemmed words*; secondly, a description of communication efforts and topics that firms opted for when dealing with sustainability. Our expected contribution is mainly methodological, as we provide suggestions regarding the advantages of performing the analysis in one of the two ways, while the research context offers insights into sustainability reporting.

1. Introduction

The use of software in research is becoming more popular due to the avalanche of information to which we are all exposed. This avalanche is composed of both scientific publications and other sources (Larsen and Von Ins, 2010; Bornmann and Mutz, 2015). Regarding the first, scholars have highlighted the science overload trend. Landhuis (2016) estimated that a new paper is uploaded every other minute on PubMed. She added that this may lead scholars to feel that they risk missing some advances and to experience the urgency to look for solutions. In the same vein, Boon (2017) observed the growth in papers publishing in the last decades. Indeed, she highlighted the milestone of 50 million papers published from 1665 to 2009, then the staggering growth that took place at a pace such that 2.5 times more scientific papers are being published every year. Therefore, the number of papers written in about 25 years will equal what had been written in the previous three centuries; the chances

are high that we will reach this peak even before then, due to the likely increase in the publishing rate, already observed to be about 4-5% in 2015 (Ware and Mabe, 2015).

Similarly, with reference to other potential sources of information, there has been enormous growth in the amount of available information due to the wide development and use of websites, blogs, and social media. Some figures are very insightful: Since the first website was launched in 1991, an almost unstoppable number of new websites has emerged, with a peak of 1.76 billion at the end of 2017 and slightly less as of October 2019, viz. 1.72 billion (Statista, 2019).

Considering the numbers above, software is needed in research, as in several cases the volume of information is unbearable for human beings. Increasingly often, if one wants to grasp meaning from extant research or from evidence based on multiple online sources, one cannot disregard software. Scholars have called for advances in the domain of software as a tool for research, as several tools are quite new and an increase in their reliability and available functions is still advocated. For instance, Guo *et al.* (2017) highlighted some limitations regarding topics set aside because of their low usage rate in customers' reviews. More recently, Moro *et al.* (2019) called for new approaches to research on papers with a focus limited to the key sections.

Consequently, this paper aims to depict the contribution that software offers in terms of performing content analysis. Studies on this analysis are not new (Krippendorff, 1980), but the progress of software, the calls for further research, and the increasing number of scholars dealing with this research approach have driven the attention focused on this topic. Indeed, several scholars are using software in their research and social media are offering new research contexts, sometimes still unexplored. Due to these factors, new research opportunities as well as methodological questions are arising and some features of software for content analysis have not yet been compared to consider their suitability. As a consequence, we propose a content analysis of the online communication of a wide set of companies dealing with sustainability, since these data would favour a comparison between two levels of investigation, namely, the *exact words* and the *stemmed words*. Particularly, software for content analysis consider every word as unique at the level named as exact words, thus separating singular from plural forms of the same word, present from past tense use of the same verb; on the contrary, software combine words as stemmed words based on the root of each word and often offering the same meaning. Stemmed words derive from a process known as lemmatisation "by which different word forms (plural, inflections, etc.) are replaced by its basic form" (Scharrow, 2017: 4).

To the best of our knowledge, Bakkalbasi and Goertzen (2014) called for this comparison, though scholars dealing with content analysis adopted one or the other, without further detailing this choice. Therefore, in the next sections we describe the theoretical framework of this research, with an overview of the automated and semi-automated techniques and tools. Then we review the use of these tools in the analysis of firms' communication. The paper continues with a description of our methodological approach and an analysis of what the 100 Most Sustainable Companies in 2019 communicate about sustainability through a social network. A discussion offers insights and the methodological implications of this research leading to stress the advantage of the different levels of analysis through automated software and the advantage in combining these results with other analysis as proximity analysis; theoretical implications as well practical implications on business communication for

sustainability are presented. At the end, we clarified limitations and suggestions for further research.

2. Theoretical framework

2.1. An overview of automated and semi-automated techniques and tools

In recent years, new technologies have provided many tools for performing both qualitative and quantitative analysis. Several instruments have been provided for text mining studies, namely, a technique to extract information through an objective perspective by using an adequate methodology. Some techniques are sentiment analysis (through tools such as Quick Search, Repustate, and Lexalytics) (Toldo, 2013; Ghiassi and Lee, 2018), semantic analysis (through tools such as Angoss and MATLAB) (Mikut and Reischl, 2011; Raut and Nathe, 2015), and content analysis (through tools such as Atlas, NVivo, and Concordance) (Neuendorf, 2016; Paulus *et al.*, 2017).

Some scholars suggest that a combination of methods or a comparison of techniques through different instruments can be more appropriate for performing text mining analysis (Toldo, 2013; Moro *et al.*, 2019), as it offers the opportunity to detect data from unstructured documents that can be from different sources (Vijayarani and Janani, 2016).

When different sources are mixed, the analysis could not be very objective, as the researcher must find a way to compare the sources and extract relevant and significant information. Automated tools allow for the converting of unstructured information such as words and sentences into quantitative and numerical values in databases (Vijayarani and Janani, 2016; Amini *et al.*, 2018) and for the implementing of iterative queries to identify and quantify textual relationships and frequencies (Jacobs, 1992; Jacobs and Rau, 1993), leading to an objective research process and undistorted (Toldo, 2013; Moro *et al.*, 2019).

About the sample, McMillan (2000) followed Krippendorff's suggestions, stating that sampling must "assure that, within the constraints imposed by available knowledge about the phenomena, each unit has the same chance of being represented in the collection of sampling units" (McMillan, 2000: 81).

Advances enable content analysts to scan "large text databases and to implement complex search queries that go beyond merely finding words and phrases (or co-occurrences of words and phrases)" (Evans, 1996: 270). Indeed, the proliferation of online text databases enables access to a wide variety of contributions, such as newspapers, magazines, official reports, and documents (Evans, 1996).

In line with our aim, a more in-depth examination of the content analysis of firms' messages intended to communicate their sustainable and responsible approaches will be performed. The choice of sustainability stems from its being a topic that addresses the interests of a variety of stakeholders. Consequently, the communication it inspires is addressed to multiple actors, leading to a wider focus in this analysis.

Thus, our first research question may be summarised as follows:

RQ1: How can exact words and stemmed words help set the analysis properly?

The next sections will describe the use of automated and semi-automated techniques in the investigation of corporate communication.

2.2. Corporate communication analysed through automated and semi-automated techniques

The analysis of reports and documents provided by both firms and institutions is performed more and more frequently by scholars aimed at investigating strategies (Bekmeier-Feuerhahn *et al.*, 2017; Werder *et al.*, 2018). With special reference to the present research, we focused our attention on the previous studies to observe the effectiveness of business communication inside and outside the company (Barnett and Loeffler, 1979).

In general, when one is talking about content analysis, Krippendorff is among the most representative scholars, providing his definition of four primary advantages of this methodology: “it is unobtrusive, it accepts unstructured material, it is context sensitive and thereby able to process symbolic forms, and it can cope with large volumes of data” (McMillan, 2000: 81). Amini *et al.* (2018: 3) highlighted the advantages of this approach, which “consists of conceptual analysis and relational analysis”.

According to our aim, an interesting contribution has been provided by Holsti (1969), who referred to the content analysis performed to observe communication strategies and processes; in detail, he identified the main aims of this approach, recalled also by McMillan (2000), namely, “to describe the characteristics of communication, to make inferences as to the antecedents of communication, and to make inferences as to the effects of communication” (McMillan, 2000: 81).

Scholars such as Kohut and Segars (1992) referred to ‘mass’ business communication vehicles as channels for spreading information among different stakeholders, referring to both periodical reports and employee publications. Haar and Kossack (1990) examined the readability and usefulness of the internal documentation of firms.

Apart from official reports and documents that inform stakeholders about business performance, more companies are using social media to communicate with their followers. In terms of business communication, Twitter is one of the most widely used social media outlets and the same applies to investigations, as recently suggested by Gomez (2020); indeed, by tweeting their followers, marketers can increase brand awareness, generate leads and revenues, foster relationships, and create brand loyalty; furthermore, social media allows companies to interact directly with their customers (Swani *et al.*, 2014: 1). In line with this, the number of Twitter users is increasing (Swani *et al.*, 2014), providing researchers with huge databases for analysing their strategies, results, vision, CSR initiatives, and so forth.

The large amount of data obliges the analysts to employ automated and semi-automated techniques and instruments, as performed in the present study.

2.3. Corporate communication on sustainability and content analysis

As mentioned above, over the last decades, even more companies are issuing reports and documents to share information about their responsible approaches and the results of their initiatives aimed at favouring sustainable development (e.g., Reilly and Hynan, 2014). The increasing amount of this kind of documentation has provided analysts with sources for performing studies on how firms communicate sustainability. An interesting example is the research conducted by Jose and Lee (2007), who carried on a content analysis of reports about

the environmental policies and practices declared by some big companies. The authors chose content analysis because they considered it to be the “primary tool used for analyzing the published information” and “a technique for making inferences by objectively and systematically identifying specified characteristics of messages” (Holsti, 1969 in Jose and Lee, 2007: 311). Indeed, such a methodology allows, among other things, the investigation of both the frequency and the connections among different concepts and words, as well as the strength of those connections through the application of specific indexes (e.g., Pearson index and Jaccard index) (Navigli and Martinelli, 2018; Miller and Howard, 2019).

Similarly, a study by Amini *et al.* (2018) offered a content analysis through Leximancer on sustainability reports published online by 500 companies belonging to nine different industries.

Another interesting contribution was offered by Landrum and Ohsowski (2018), whose study focused on the analysis of sustainability or CSR reports, aiming to “understand corporate worldviews regarding the meaning of sustainability or CSR” (Landrum and Ohsowski, 2018: 128), through a content analysis, recalling previous applications with similar goals (e.g., Manetti and Toccafondi, 2014; Lock and Seele, 2016; Landrum and Ohsowski, 2018). In detail, the authors analysed the frequency of keywords among the different stages of corporate sustainability as content categories proposed by Landrum (2015) himself. Calls for research on sustainability through content analysis were addressed by scholars addressing the need of overcoming data selection bias (Mani *et al.*, 2018) as well as dealing with wider datasets (Tsalis *et al.*, 2020)

Therefore, our second research question is as follows:

RQ2: What are the key topics in communicating a firm’s approach to sustainability? How are these topics communicated?

3. Research aim

In line with the aforementioned calls for research (e.g., Bakkalbasi and Goertzen, 2014; Mani *et al.*, 2018; Tsalis *et al.*, 2020), we investigate the content of the communication on sustainability with a two-fold aim: firstly, a comparison between *exact words* and *stemmed words*; secondly, a description of communication efforts and topics that firms opted for when dealing with sustainability. Thus, the methodological contribution would provide suggestions about the advantages of performing the analysis in one of the two ways, while the research context offers insights into sustainability reporting.

The structure of our aim and the aforementioned RQs led us to opt for qualitative and exploratory research (Johnston *et al.*, 1999). Krippendorff (2004) set the standard for performing content analyses aimed at being objective while inferring data. Three standards describe a suitable approach to this analysis, namely, identifications, evaluations, and judgments. More in detail, the first provides an overall description of the phenomena investigated, while the evaluation represents the research stage in which the analysts grasp meaning from words and consider their usage frequency and the contextual meaning. Finally, to make the analysis useful for advances, the evidence should be reconducted to general theories, models, and key concepts.

4. Research methodology

We opted for a qualitative content analysis through NVivo, as this software has been recognised as one of the most reliable; this approach also retained our ability to keep the analysis at an objective level, as proposed by Bergman (2010). Indeed, in his book, Bergman stated that NVivo represents a proper instrument for meeting the needs of objective analysis as advocated by scholars even in the 1950s (e.g., Berelson, 1952) while preserving the chance to simplify and organise data, as well as systematise the evidence, even in dealing with a huge amount of sources. This is also due to the focus on the most frequent words – usually the 50 most cited words (e.g., Descubes *et al.*, 2018) – to limit the analysis to the key evidence.

Due to our focus on communicating sustainability, our unit of analysis is a single message conveyed by a firm. According to Krippendorff (2008), each unit of analysis should be distinguished from any other data. Thus, every message posted online by a firm may mirror this feature. Consequently, this criterion led data collection together with the choice of our research context. The context of analysis consists of the 2019 100 Most Sustainable Firms according to the report Corporate Knights as published on Forbes (2020). This choice stems from the inspiring role these firms may play in the domain of sustainable actions in the business domain and from the fact that their communication may potentially lead to guidelines for other businesses. Therefore, data collection took place with reference to the messages of these firms. We selected Twitter as a context for data collection due to the opportunities to focus on homogeneous messages in terms of length and to clearly distinguish between what the firm posted and what other entities can write in a social media channel; additionally, our selection stemmed from the fact that previous studies have considered Twitter to be a context for communicating about the social issues that a firm addresses (Araujo and Kollat, 2018) and engaging stakeholders in achieving goals that a firm considers valuable (Mmic and Alamaraz, 2013).

The sources have been collected through the add-on NCapture for NVivo (version 12.0), favouring unbiased data collection. Data collection took place at the beginning of 2020 and the tweets were published in a time span from late 2015 to January 2020. Tweets were downloaded for 84 of the 100 firms; the remaining 16 business profiles were inactive or unavailable, or contained messages in languages other than English. No messages were discarded for not being focused on sustainability.

Finally, the research process is set as a qualitative research, albeit quantitative data are offered by the software to describe frequencies; moreover, indices and coefficients are used – as it will be further described – but the essence of cluster analysis remains qualitative. Indeed, our research process is articulated on four levels, each of which is complemented by our interpretation of data:

- word frequency of the *exact words*;
- word frequency of the *stemmed words*, leading to a comparison with the previous level of analysis;
- choice of the most suitable result between the previous levels and a qualitative cluster analysis of it;
- further focus on key results through proximity analysis.

Data collection offered different approaches adopted by companies in using Twitter to communicate the essence of their activities. Indeed, 2 out of the 84 firms considered in this

research posted fewer than 100 tweets, while most of the others tweeted, on average, about 10,000 times. In total, we collected and analysed 876,190 tweets. To ensure reliability, the first two levels of the research process described earlier were performed several times with sub-samples. Negligible differences emerged among the sub-samples. Therefore, we decided to consider the entire dataset shaped by 84 firms along the four levels of analysis.

Finally, the reliability of data is further ensured by the evidence of the qualitative cluster analysis, supported by Jaccard's similarity coefficient set at +1 on a range from -1 to +1. This criterion allows us to highlight the similarities among the sources, namely, the similar usage of words in combination. Thus, the index set at 1.0 represents the highest level of similarity in the messages addressed by the 84 firms, leading to consideration of the most insightful and common combinations among topics.

5. Findings

The analysis consists of the four levels mentioned earlier and starts with the two 'word frequencies'—one based on *exact words* and the other based on *stemmed words*. These two analyses are described together in the following lines, as their comparison would lead to the identification of the one most suitable for highlighting the key topics in firms' communication about sustainability.

The comparison led us to consider the *stemmed words* analysis as more suitable for this research, because of the highlighting of topics related to the debate on sustainability and because the combining of some results positively affects the emergence of pivotal themes in this research domain. Indeed, sustainability – together with connected words such as sustainable and sustain, and combined forms such as sustainability-based and sustainability-driven – moved from the 42nd position in the ranking of the most frequently cited words in the *exact words* analysis to the 5th position; similarly, innovation is a key topic in describing what firms do to favour sustainability-oriented intervention. The group of terms related to innovation ranked 6th in the *stemmed words* analysis, while it was 45th in the other analysis; this change stems from the use of terms such as innovate, innovative, innovation-driven, innovating, and so on. A theme strictly connected to both sustainability and innovation is technology, and even in this case, the *stemmed words* analysis offers more relevance, moving to the 13th position from the 24th position in the *exact words* ranking. Additional examples leading us to choose the second kind of analysis are sharing (rising to 26th from 78th), develop (rising to 33rd from 88th), and design (rising to 38th from 57th). These three examples are particularly relevant because it is common to limit the analysis to the 50 top-cited words, as addressed before. Therefore, these topics would have been set aside, despite their relevance to the debate on sustainability. The limit of 50 words has also been considered to check for any potential 'dark side' in using the second words ranking; indeed, no relevant themes moved out of the focus during shifting from the *exact words* analysis to the *stemmed words* analysis; as a consequence, the analysis retained its objectivity, as while we chose to opt for one of the two rankings, no key topics were lost. Additionally, it is interesting to stress that some topics emerged in the second typology of analysis, though they were not among the top 100 most-cited words. This occurred with product (grouped with production, produced, and so on), which ranked 40th in the second analysis.

As a consequence, the analysis continued along the third and fourth levels based on the evidence acquired through the stemmed words. For this reason, Table 5.1 describes the ranking of the most-cited words, aimed at offering a first overview of the results regarding the content of the most sustainable firms' communication through Twitter.

At first sight, the most frequently used words suggest some actions in which firms are mainly involved, as is the case with *help*, *learn*, *innovate*, *read*, *find*, and *make*. All these verbs introduce the activities that firms are promoting in their online communication. Before continuing with the qualitative cluster analysis, it is relevant to stress that these results are often related to interactions with other actors, such as helping someone (e.g., the customers), learning from someone (e.g., partners or competitors), and *reading* as a way to stimulate various actors to read what the firm did or proposes.

Apart from these basic considerations, the key part of the analysis deals with the qualitative cluster analysis; this latter offers the opportunity to identify the ties among topics. Thus, it is possible to grasp meanings from the use of terms together with others.

Word	Count	Similar Words	Word	Count	Similar Words	Word	Count	Similar Words
new	12904	#new, @new, new, new, 'new	living	5220	#live, #living, @living, live, 'live', live', lived, lived', lively, lives, lives', living, 'living, living'	teams	4110	#team, #teams, @team, team, team', 'team, team', teamed, teaming, teams, teams'
helps	8345	help, help', helped, helpful, helping, helps	see	5142	see, see, see', seeing, seeing', sees	company	4086	#companies, #company, companies, companies', companies', company, company', company', companys, companys'
learn	8242	#learn, #learning, learn, learn, 'learn, learned, learning, learning', 'learning', learnings, learns	future	5088	#futur, #future, #futures, @future, @futurism, futur, futurability, future, future', 'future, future', futures, futuric, futurism, futurs	timing	4084	#time, @time, time, time', time', timed, timely, times, timing, timings
year	7487	#year, year, year', year', year'at, yearly, years, years, years', year's#nationalanthem, years'	now	4893	#now, now, now	design	4049	#design, #designed, #designer, #designers, #designs, @design, 'design, design, design', designated, designating, designation, designations, designed, designer, designers, designers', designing, designs
sustain*	6997	#sustainability, #sustainable, #sustainably, @sustainability, @sustainly, sustain, sustainability, sustainability', 'sustainability, sustainability', sustainable, 'sustainable, sustainable', sustainably, sustained, sustainers, sustainability, sustaining, sustains	ones	4884	#one, one, one', 'one, one', ones	via	4037	via
innov*	6858	#innovability, #innovants, #innovate, #innovating, #innovation, #innovation#ucb90, #innovations, #innovative, #innovator, #innovators, #innovor, @innov, @innovation, innov, innovability, innovant, innovants, innovate, innovated, innovates, innovating, innovation, innovation, innovation', innovation', innovations, innovations', innovations', innovative, innovative', innovativeness, innovator, innovators, innove, innovent, innovor	energy	4745	#energi, #energie, #energie#processdeveloper, #energies, #energy, @energiency, @energy, energi, energie, energies, energies', energy, energy', 'energy, energy'	products	4003	#product, #production, #productivity, #products, @product, product, product', product'sourcing, product', production, productions, productive, productively, productivity, products, products', products'
today	6844	#today, @today, today, todays, todays'	using	4727	use, use', used, useful, useful', usefulness, uses, using	needs	3981	need, needed, needing, needs
thanks	6833	#thankful, #thanks, thank, thanked, thankful, thankfully, thanking, thanks, @thanks, thanks#cgceco, thanks'	solutions	4703	#solution, #solutions, @solution, solutely, solution, solution', solutions, solutions', solutions'	first	3963	#first, @first, first, first, first#5g, first', firstly, firstlys
world	6620	#world, @world, world, world', 'world, world#brandstorm2017#monthlydigest, world', world's2nd, worlds, worlds', worlds'	sharing	4630	#share, #shared, #shares, #sharing, @shareing, 'share, share, share, shared, shares, sharing, sharing, sharing#	people	3953	#people, @people, people, people', 'people, people', peoples, peoples', peoples'
works	6290	#work, #working, work, work', 'work, work', worked, working, 'working, workings, works, works'	industry	4568	#industrial, #industrie, #industries, #industry, industri, industrial, 'industrial, industrialization, industrialize, industrialized, industrializing, industrials, industrie, industries, industries', 'industries, industry, industry', industry@	changing	3896	#change, #changes, @changing, chang, chang'an, change, change', 'change, change#, change#cwmyc, change', changed, changes, changing, 'changing
read	6019	#read, #reading, read, read, 'read, reading, readings, reads	taking	4527	take, takes, taking	waters	3858	#water, #water'courses, #water'efficiency, @water, water, watered, watering, waters
find	5779	find, find, finde, findes, finding, findings, finds	joins	4515	@join, join, joine, joined, joining, 'joining, joins	powers	3763	#power, #powerful, #powering, #powers, @power, power, power', powered, powered, powerful, powering, powers
technology	5772	#technological, #technologically, #technologies, #technology, @technology, technolog, technological, technologically, technologies, technologies', technologies', technology, 'technology, technology'	global	4474	#global, #gloaled, #globalization, @global, global, global, 'global, globale, globales, globalism, globalization, globalized, globalizing, globally	digits	3753	#digital, #digitalization, #digitalized, #digitization, #digitize, #digitized, #digitizing, @digital, digit, digital, digital, 'digital, digitalization, digitalize, digitalized, digitalizes, digitalizing, digitally, digitization, digitize, digitized, digitizing, digits
busy	5671	#business, #businesses, @business, @business', business, business', 'business, businesses, businesses', businesses', busy, busying	ceos	4284	#ceo, #ceos, @ceo, ceo, ceo', ceos	check	3590	#check, #checking, check, check, checked, checking, checks
making	5477	@make, make, 'make, makes, making, 'making	supports	4242	#support, support, supported, supported, supporter, supporters, supporting, 'supporting, supportive, supports	customs	3554	#custom, #customer, #customers, #customers', #customization, #customize, custom, customer, 'customer, customers, customers', customers', customization, customize, customized, customizes, customizing, customs
looks	5421	look, looked, looking, looking, looks	develops	4173	#developer, #developers, #developing, #development, #development', #developments, develop, developed, developement, developer, developers, developers', developing, 'developing, development, development', development', developments, developed, developing, develops	projects	3484	#project, #projection, #projects, project, projected, projecting, projection, projections, projects, projects'
day	5364	#day, day, day', 'day, day', days	getting	4159	@get, 'getting, get gets, getting, 'getting			

Table 5.1. Top 50 most-cited words (stemmed).

A multitude of ties emerged from the analysis, with some offering interesting insights; by the way, this does not mean the results with no ties have to be disregarded, because we set the highest possible level of significance. Other results were quite expected and, therefore, won't be further processed. With an aim to clarify what the most sustainable firms communicated, the ties that emerged are depicted in Figure 5.1.

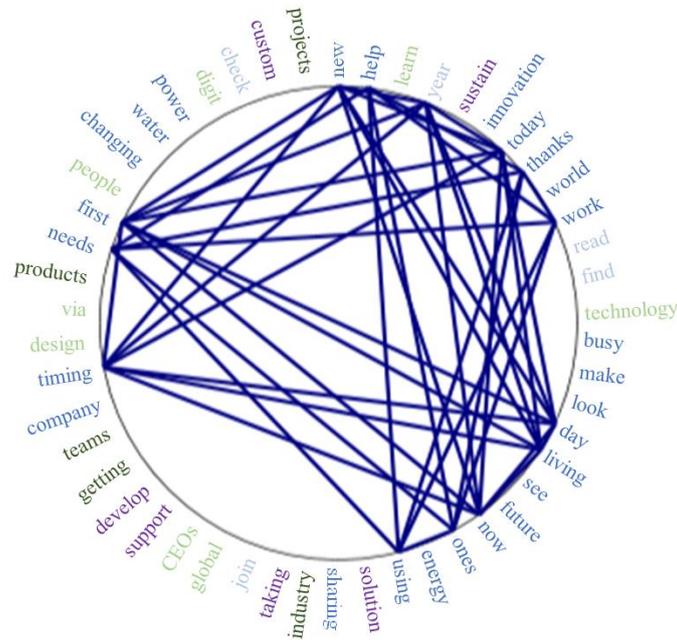


Figure 5.1. Qualitative cluster analysis of the 50 top-cited stemmed words (Jaccard's index = 1).

The first tie analysed in detail consists of *new* and *living*; these two words describe how some actions and actors may affect the lives of others, as in the efforts that firms make to ensure liveable conditions for local inhabitants or preserve biodiversity. Regarding the first, Generali designed a project in cooperation with the Human Safety Network to improve the living conditions of refugees and contribute to their development in the local economy by offering support in starting new businesses. The message is properly summarised as giving people the chance to “take back their dignity”. Concrete efforts are also addressed by the Spanish firm Acciona; one of its messages stressed how its experts may be supportive in designing a better planet and improving the transport network. In this way, the firm expects to cooperate at a local level to promote safer conditions in managing local congestion, ensuring local jobs, and leveraging new technological capacities. The energy provider Enel expanded the vision of what to do to ensure a better future, as it called for a transformation to a future in which the way people live, the role of economics, and the use of natural resources may be directed towards several challenges. Indeed, the firm identified the need to invest in technology and to partner with other big companies to fight global issues such as climate change. Similarly, the US company McCormick Corporation pays attention to the needs of local communities in assessing its global sourcing practices; to that end, the firm's experts visit farming cooperatives in remote communities to learn about new ways to obtain more social and economic benefits for inhabitants.

The second tie that was investigated in more detail is between *help* and *work*. In most cases, cooperation and involvement are the two main goals that these words depict, as firms are either describing or seeking solutions that allow them to work with other actors in pursuit of a better future. Such involvement is mirrored in communication that shares proposals and firms' efforts with a multitude of actors. A message from Ørsted, a Danish company offering green energy solutions, asked for help in sharing a short movie in which the firm showed how it works to achieve a greener world. The idea of working together to help someone else is also proposed by Cisco, as the firm describes how everyone can help in supporting the lives of other people, especially the homeless, by sharing experiences and giving away for charity. Additionally, *work* is a term used to explicitly refer to working conditions. Indeed, firms do their best to help people experience better conditions at work. This is mirrored in a message, again from Cisco, telling a story about a worker who faced gender-based discrimination and who now offers support to others with similar problems. Efforts aimed at counteracting discrimination at work are also reported by Hewlett Packard in the form of a program that helps people with autism have a chance to work, even in companies other than Hewlett Packard.

One more tie offering interesting insights derives from the combination of *year* and *living*; with *year* being related to time-framing efforts already performed, in progress, or planned, this tie offers a way to frequently report on actions or projects. The life science company Sanofi reported on the acknowledgement of its efforts in a report issued by Dow Jones on sustainable firms, recognising a decade of actions towards promoting a socially responsible vision of its activities. Similarly, Siemens celebrated being ranked as the best firm in its industry in terms of sustainability; the firm stated that throughout the previous year, it served society without sacrificing its sustainability goals. Additionally, the pharmaceutical firm Takeda reported on its previous-year results and expressed that it was proud of having exceeded the aimed targets, thus enhancing sustainability at both the firm and supply chain levels. Additionally, concrete results are the main goal for Dassault Systèmes, a French software company that launched a challenge to design innovative solutions leveraging data management to solve societal and environmental issues. Other results, though related to internal issues, are reported by Standard Bank Group, as the firm tracks efforts performed throughout the year thanks to a partnership aimed at defining gender equity in the workplace. Additionally, the firm calls for participation to advance these results in the upcoming year.

The focus on the fourth tie stresses internal issues related to sustainability; this tie depends on the co-occurrence of *works* and *using*. In any event, reporting on the sustainability of internal operations is often partnered with the assessment of the consequences on the outside world. Therefore, it is common to identify examples of firms establishing green activities, based on lower levels of resource usage, to work for the environment, while continuing their business. The DSM company implemented the principles of circular economy in its daily operations to reuse and recycle, thus stressing that what it does is aligned with its environmental-friendly vision and sets best practices. Similarly, actions addressing internal issues do not limit the contribution of external actors; indeed, Schneider Electric paid a great amount of attention to efficiency. It is seen as an internal criterion for evaluating operations as well as innovation, and also as a philosophy to be spread among partners, aimed at creating sustainable supply chains. In this way, the firm expands the beneficial effects of a wise resource management system and of process supported by awareness in the evaluation of environmental impacts. Similar evidence is offered by Autodesk, as sustainable solutions are not only at the core of

their offerings but also represent a way to use traditional methods without forgetting about the goal of working for better environmental conditions. A parallel between internal objectives and external impacts is offered by Outotec, which implemented various new technologies into its processes and compared the performance with that of its competitors, while considering the emissions avoided and the beneficial effects in counteracting climate change. Thus, Outotec's sustainability reporting is a way to both self-evaluate internal results and assess the beneficial impacts on the environment.

The focus on the ties offered through the qualitative cluster analysis led to the consideration of key issues and approaches in firms' communication about sustainability. When one is dealing with a large amount of data, it can be useful to go more in-depth in the analysis and look for specific topics. This action can be carried out at any time during the research, though using it as a way to go more in-depth after a qualitative cluster analysis may offer valuable insights. Indeed, the qualitative cluster analysis above has not offered statistically significant ties based on Jaccard's index; thus, key themes in this research, such as sustainability, innovation, and technology, to mention just a few, may be further detailed. Therefore, the proximity analyses of these words are offered next. Sustainability is at the core of communication because it mirrors several elements, namely, a vision, an aim to be shared, and a way to assess a firm's efforts. Chr. Hansen, a bioscience firm, offers an example of this multi-fold usage of the word *sustainability*; indeed, in a message, the firm reported being inspired at an international forum regarding how sustainability should be the not-to-be-missed philosophy for the future, as it represents a lens for setting goals to be achieved throughout the supply chain and industry. In another message, Chr. Hansen stated that it is proud of being highly-ranked among sustainable firms, parallel to being considered one of the best food ingredients producer in the world. Therefore, sustainable goals are compatible with quality standards and customer satisfaction.

One more theme we analysed through a proximity analysis was innovation, described as both a way of thinking and a way of acting, especially when other actors are involved in the challenges brought about by sustainability. Thus, innovation is considered a path to designing more sustainable operations and achieving better conditions for the environment. Metso Group described its efforts to promote sustainability, driven by innovation through a challenge and an award; indeed, internal teams were created and invited to propose innovative solutions to improve firms' offerings, considering investments as a way to increase both productivity and environmental benefits. Panasonic enlarged this vision and stimulated collaboration to address sustainability challenges through open innovation, thereby inviting partners to focus on the Sustainable Development Goals set by the United Nations and implementing them along the road to a company's success. L'Oréal, too, regarded innovation as permeating internal operations and communications; indeed, the firm considers innovation as an opportunity to perform better and make a positive impact. This two-fold aim should inspire the firm's everyday activities.

The third and final theme to which we devoted attention is technology, due to the frequency of its use and its instrumental role in innovating for sustainability. The proximity analysis focuses on technology as an internal tool and aggregator; indeed, technology is the essence of several innovations. Thus, it is implemented in firms as a new instrument to advance new solutions. At the same time, technology is seen as a way to group actors into activities and projects run by multiple actors in pursuit of sustainable results, as well as the essence of some communication channels. Accenture offers an example of technology regarded as an internal

tool in questioning the approach to sustainability. The firm stated that technology and business models are no longer aligned with people's wants and environmental-based needs. Therefore, technology should be adjusted so that it can be used to align a firm's operations with global needs. Valeo adopted a similar perspective but expanded it in considering technology as a key element of its offerings, as well as a way to involve users and partners in the challenge to pursue sustainability. Indeed, the firm partners the word *technology* with *change*, referencing the change that technology can bring to actors' ways of doing things.

6. Discussion

In line with calls for research (e.g., Bakkalbasi and Goertzen, 2014; Mani *et al.*, 2018; Tsalis *et al.*, 2020), we aimed to accomplish our study's goals by performing content analysis, as it offers the advantage of focusing on common issues. Thus, the representativeness of the results is supported by the analysis itself, while the examples we provided in the second part clarify and provide a more concrete understanding of some issues.

In detail, we investigated the content of the communication about sustainability through the messages that companies tweeted. We opted to carry out the analysis of word frequency by adopting two approaches, to observe how insights change if one takes into account the 'exact words' or the 'stemmed words' (Jacobs, 1992; Jacobs and Rau, 1993), which is useful for verifying any connection among groups of words with the same root, and which is in line with the suggestions of Landrum and Ohsowski (2018).

One of the most interesting results is related to the relevance acquired by the concepts of sustainability and innovation; indeed, to perform the analysis, we collected all the tweets of the top 100 sustainable companies, aiming to avoid any bias. Despite this, the results show that sustainability was the 5th highest-cited in the ranking, which innovation was 6th, behind less significant terms such as *new*, *helps*, *learn*, and *year*. This shows how significant these issues are to the companies, at least in their communication to stakeholders—in some cases, even more than their core business. Furthermore, it must be noted that most of the tweets were quite old and were not influenced by Forbes' research. Therefore, in general, sustainability is at the core of communication because it mirrors several elements, namely, a vision, an aim to be shared, and a way to assess a firm's efforts.

7. Implications

7.1. Methodological implications

As discussed in the literature review, new technologies are providing researchers with a huge number of sources—especially online sources—whose analysis would be extremely complex without the use of a quantitative method (Jacobs and Rau, 1993; Vijayarani and Janani, 2016). This is the primary reason why we opted for a content analysis through NVivo. Our proposed methodological approach aimed to show the significance of different methodological tools in performing analyses of specific topics.

The results of the analysis performed on the three main topics underline that the frequency of the exact words does not always indicate the real relevance of a topic, as they can be

transversal in different themes; on the other hand, when there are no recurring combinations, the analysis of stemmed words cannot reveal significant results.

Answering RQ1, we can state that, in general, our analysis shows that stemmed words analysis is more suitable for this research, as it allows us to highlight the most relevant topics in the debate on sustainability, thereby favouring the emergence of pivotal themes in this research domain.

Furthermore, the results show that if our analysis had been focused on pre-assumed topics such as sustainability, innovation, environment, and green (Manetti and Toccafondi, 2014; Lock and Seele, 2016; Landrum and Ohsowski, 2018), the results would have been quite expected and, in any case, not widespread among all companies. Meanwhile, the qualitative cluster analysis shows unexpected words, even if they are central to the companies' communication efforts.

In line with this, our approach can be considered a more objective approach. Through the double analysis and the comparison, we avoided any subjective bias thanks to the qualitative cluster analysis, which is based on an objective evaluation using software with a statistical setting (Amini *et al.*, 2018).

As discussed in the literature review, automated tools allow for the conversion of unstructured information—such as words or sentences—into quantitative values to be collected in databases (Jacobs, 1992; Jacobs and Rau, 1993; Vijayarani and Janani, 2016), leading to an objective research process and results that are not affected by biases (Vijayarani and Janani, 2016; Amini *et al.*, 2018). It must be noted that we are not suggesting that stemmed words analysis should always be preferred; rather, we are suggesting a comparison of the two approaches when confirmation of the emerging results is needed.

In addition, we suggest running a proximity analysis after the qualitative cluster analysis to focus on specific issues. This is particularly useful with a huge amount of data or to go into the analysis in more detail and overcome the limits of qualitative cluster analysis, namely, the consideration of a couple of words. Through proximity, it is possible to expand the analysis to groups of words. In addition, performing a proximity analysis after a qualitative cluster analysis helps to better contextualise the emerging link. This allows one to keep the objective approach as the focus is on the most recurring themes and not on topics chosen by the authors.

7.2. Theoretical implications (for sustainability)

We focused our research on how firms communicate their sustainable approach. In this regard, the writing style adopted by companies highlights that they pay attention mainly to how—as compared to what—to communicate

As underlined in the discussion section, sustainability appears to be one of the most discussed topics, even if this can be considered an expected result, as we analysed companies that Forbes indicated were the most sustainability-oriented ones.

Regarding RQ2, the findings showed that, in their communication process, companies have a double aim: (1) to report the results of their activities, with special reference to those that have a responsible and sustainable approach (Reilly and Hynan, 2014; Landrum and Ohsowski, 2018); and (2) to involve all stakeholders, encouraging them to adopt the same responsible approach towards sustainable development (Landrum and Ohsowski, 2018).

When discussing sustainability issues, companies tend to combine the three different approaches towards sustainability, namely, the social, environmental, and economic spheres, as they follow the most relevant international and global trends (Manetti and Toccafondi, 2014; Lock and Seele, 2016; Landrum and Ohsowski, 2018). Also, social and environmental issues are more public-oriented while economic sustainability is pursued mainly to make the business process more efficient.

Apart from sustainability, the most considered topics are innovation and technology, together with innovation-based and technology-based. Innovation emerged as the 6th ranked in the stemmed words list and is described both as a way of thinking and of acting, as companies state that they innovate to make their processes more efficient and favour sustainable development. On the other hand, technology emerges as a tool that the company can use to implement and develop innovative initiatives and as an aggregator to carry out the communication process.

7.3. Practical implications

The companies we analysed provide suggestions regarding how to communicate the sustainability-oriented corporate approach to both attract customers' interest and involve all the stakeholders, encouraging them to adopt the same approach (Reilly and Hynan, 2014; Landrum and Ohsowski, 2018). As mentioned in the theoretical implications, when discussing sustainability, companies consider all three spheres, even if social and environmental issues are usually more explicit. Indeed, the most frequently used words—*help, learn, innovate, read, find, make, and look*—suggest that companies become involved in activities to improve the quality of life of the local community as well as of society in general. Because Forbes considered these 100 companies to be the most sustainability-oriented ones, we can assume that their communication approach is successful. Therefore, they can be considered as engaging in best practices to inspire companies in communicating a responsible approach towards stakeholders and the community at large. This statement not only refers to the content of the communication but, mostly, to the way companies communicate. The interest stimulated shows the strength of both the communication tool adopted (a widely used social media platform such as Twitter) and the issues discussed, which are among the hottest topics at a global level.

8. Conclusion

In line with this call for research, our study aimed to observe the usefulness of automated and semi-automated techniques for managing a large amount of data and information.

To accomplish this goal, we investigated how companies communicate about sustainability. In detail, we collected all the tweets posted by the 100 companies that Forbes (2020) considered the most sustainable. The large amount of data required the use of software allowing us to perform the analysis while maintaining an objective approach. The results showed that automated tools, such as the software we used, ensure an objective interpretation of the results. However, in some cases, a more in-depth analysis through a subjective perspective can help the researcher better understand the available information. In this regard, a further study on reports provided by companies through a narrative approach can offer more insight into the

approach they adopted and their aim when communicating. Furthermore, it would be interesting to repeat the analysis on users' comments to determine how they perceive companies' messages and initiatives.

Keywords

sustainability; content analysis; Twitter; corporate communication; qualitative analysis

Reference list

- Amini, M., Bienstock, C.C., and Narcum, J.A. (2018), "Status of Corporate Sustainability: A Content Analysis of Fortune 500 Companies", *Business Strategy and the Environment*, 27 (8): 1450–1461.
- Araujo, T., and Kollat, J. (2018), "Communicating effectively about CSR on Twitter. *Internet Research*, 28(2), 419–431.
- Bakkalbasi, N., and Goertzen, M. (2014), "Scholarly E-Book across Disciplines: Content Analysis of Usage Reports and Search Terms", *CUNY Library Assessment Conference*. Available online at https://academicworks.cuny.edu/ols_proceedings_lac/4 (last accessed: February 20, 2020).
- Barnett, A., and Leoffler, K. (1979), "Readability of Accounting and Auditing Messages", *The Journal of Business Communication* (1973), 16 (3): 49–59.
- Bekmeier-Feuerhahn, S., Bögel, P.M., and Koch, C. (2017), "Investigating Internal CSR Communication: Building a Theoretical Framework", in S. Diehl *et al.* (eds), *Handbook of Integrated CSR Communication*, 89–107, Cham, Switzerland: Springer.
- Berelson, B. (1952), *Content Analysis in Communication Research*, New York, NY: Free Press.
- Bergman, M.M. (2010), "Hermeneutic Content Analysis: Textual and Audiovisual Analyses within a Mixed Methods Framework", in A. Tashakkori and C. Teddlie (eds), *SAGE Handbook of Mixed Methods in Social and Behavioural Research* (2nd edn.), 385–394, Thousand Oaks, CA: SAGE.
- Boon, S. (2017), "21st Century Science Overload", *CDN Science Pub*. Available online at <http://blog.cdnsiencepub.com/21st-century-science-overload/> (last accessed: February 20, 2020).
- Bornmann, L., and Mutz, R. (2015), "Growth Rates of Modern Science: A Bibliometric Analysis Based on the Number of Publications and Cited References", *Journal of the Association for Information Science and Technology*, 66 (11): 2215–2222.
- Descubes, I., McNamara, T., and Claasen, C. (2018), "E-Marketing Communications of Trophy Hunting Providers in Namibia: Evidence of Ethics and Fairness in an Apparently Unethical and Unfair Industry?", *Current Issues in Tourism*, 21 (12): 1349–1354.
- Evans, W. (1996), "Computer-Supported Content Analysis: Trends, Tools, and Techniques", *Social Science Computer Review*, 14 (3): 269–279.
- Forbes (2020), "The Most Sustainable Companies in 2019", *Forbes*. Available online at <https://www.forbes.com/sites/karstenstrauss/2019/01/22/the-most-sustainable-companies-in-2019/#151762136d7d> (last accessed: February 20, 2020).

- Ghiassi, M., and Lee, S. (2018), "A Domain Transferable Lexicon Set for Twitter Sentiment Analysis Using a Supervised Machine Learning Approach", *Expert Systems with Applications*, 106: 197–216.
- Gomez, L.M. (2020), "The State of Social Media Research in CSR Communication", in D. Crowther and S. Seifi (eds), *The Palgrave Handbook of Corporate Social Responsibility*, 1–23, Cham, Switzerland: Springer.
- Guo, Y., Barnes, S.J., and Jia, Q. (2017), "Mining Meaning from Online Ratings and Reviews: Tourist Satisfaction Analysis Using Latent Dirichlet Allocation", *Tourism Management*, 59: 467–483.
- Haar, J., and Kossack, S. (1990), "Employee Benefit Packages: How Understandable Are They?", *The Journal of Business Communication (1973)*, 27 (2): 185–200.
- Holsti, O.R. (1969), *Content Analysis for the Social Sciences and Humanities*, Reading, MA: Addison-Wesley.
- Jacobs, P.S. (1992), "TRUMP: A Transportable Language Understanding Program", *International Journal of Intelligent Systems*, 7 (3): 245–276.
- Jacobs, P.S., and Rau, L.F. (1993), "Innovations in Text Interpretation", *Artificial Intelligence*, 63 (1-2): 143–191.
- Johnston, W.J., Leach, M.P., and Liu, A.H. (1999), "Theory Testing Using Case Studies in Business-to-Business Research", *Industrial Marketing Management*, 28 (3): 201–213.
- Jose, A., and Lee, S.M. (2007), "Environmental Reporting of Global Corporations: A Content Analysis Based on Website Disclosures", *Journal of Business Ethics*, 72 (4): 307–321.
- Kohut, G.F., and Segars, A.H. (1992), "The President's Letter to Stockholders: An Examination of Corporate Communication Strategy", *The Journal of Business Communication (1973)*, 29 (1): 7–21.
- Krippendorff, K. (1980), "Validity in Content Analysis", in E. Mochmann (ed.), *Computerstrategien für die kommunikationsanalyse*, 69–112, Frankfurt: Campus.
- Krippendorff, K. (2004), "Reliability in Content Analysis: Some Common Misconceptions and Recommendations", *Human Communication Research*, 30 (3): 411–433.
- Krippendorff, K. (2008), "Systematic and Random Disagreement and the Reliability of Nominal Data", *Communication Methods and Measures*, 2 (4): 323–338.
- Landhuis, E. (2016), "Scientific Literature: Information Overload", *Nature*, 535 (7612): 457–458.
- Landrum, N. (2015), "Evolution of corporate sustainability along the sustainability spectrum. Poster presented at the Loyola University Chicago Climate Change Conference, Chicago, IL (Vol. 20). Available online at <https://www.luc.edu/media/lucedu/sustainability-new/images420/climatechange/postersccc2015/Evolution%20of%20Corporate%20Sustainability%20Along%20the%20Sustainability%20Spectrum.pdf> (last accessed: February 20, 2020).
- Landrum, N.E., and Ohsowski, B. (2018), "Identifying Worldviews on Corporate Sustainability: A Content Analysis of Corporate Sustainability Reports", *Business Strategy and the Environment*, 27 (1): 128–151.
- Larsen, P., and Von Ins, M. (2010), "The Rate of Growth in Scientific Publication and the Decline in Coverage Provided by Science Citation Index", *Scientometrics*, 84 (3): 575–603.

- Mamic, L.I., and Almaraz, I.A. (2013), "How the Larger Corporations Engage with Stakeholders through Twitter", *International Journal of Market Research*, 55 (6): 851–872.
- Mani, V., Gunasekaran, A., and Delgado, C. (2018), "Supply Chain Social Sustainability: Standard Adoption Practices in Portuguese Manufacturing Firms", *International Journal of Production Economics*, 198: 149–164.
- McMillan, S.J. (2000), "The Microscope and the Moving Target: The Challenge of Applying Content Analysis to the World Wide Web", *Journalism & Mass Communication Quarterly*, 77 (1): 80–98.
- Mikut, R., and Reischl, M. (2011), "Data Mining Tools", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1 (5): 431–443.
- Miller, A.H., and Howard, A. (2019), "Tax Strategies Employed by Overseas American Individuals and SMEs", *The Journal of Social Sciences Research*, 5 (3): 690–699.
- Moro, S., Pires, G., Rita, P., and Cortez, P. (2019), "A Text Mining and Topic Modelling Perspective of Ethnic Marketing Research", *Journal of Business Research*, 103: 275–285.
- Navigli, R., and Martelli, F. (2019), "An Overview of Word and Sense Similarity", *Natural Language Engineering*, 25 (6): 693–714.
- Neuendorf, K.A. (2016), *The Content Analysis Guidebook*, Thousand Oaks, CA: Sage.
- Paulus, T., Woods, M., Atkins, D.P., and Macklin, R. (2017), "The Discourse of QDAS: Reporting Practices of ATLAS.ti and NVivo Users with Implications for Best Practices", *International Journal of Social Research Methodology*, 20 (1): 35–47.
- Raut, R., and Nathe, A. (2015), "Comparative Study of Commercial Data Mining Tools. International Journal of Electronics", *Communication and Soft Computing Science & Engineering (IJECSCE)*, 8: 128–132.
- Reilly, A.H., and Hynan, K.A. (2014), "Corporate Communication, Sustainability, and Social Media: It's Not Easy (Really) Being Green", *Business Horizons*, 57 (6): 747–758.
- Scharkow, M. (2017), "Content Analysis, Automatic", in J. Matthes *et al.* (eds), *The International Encyclopedia of Communication Research Methods*, 1–14, Hoboken, NJ: John Wiley & Sons.
- Statista (2019, October 28), "How Many Websites Are There?" [Blog entry], *Statista*. Available online at <https://www.statista.com/chart/19058/how-many-websites-are-there/> (last accessed: February 20, 2020).
- Swani, K., Brown, B.P., and Milne, G.R. (2014), "Should Tweets Differ for B2B and B2C? An Analysis of Fortune 500 Companies' Twitter Communications", *Industrial Marketing Management*, 43 (5): 873–881.
- Toldo, L. (2013), "Text Mining Fundamentals", in J. Liebowitz (ed.), *Business Analytics: An Introduction*, 149–180, Boca Raton, FL: CRC Press.
- Tsalis, T.A., Malamateniou, K.E., Koulouriotis, D., and Nikolaou, I.E. (2020), "New Challenges for Corporate Sustainability Reporting: United Nations' 2030 Agenda for Sustainable Development and the Sustainable Development Goals", *Corporate Social Responsibility and Environmental Management*, 27 (4): 1617–1629.
- Vijayarani, S., and Janani, R. (2016), "Text Mining: Open Source Tokenization Tools – An Analysis", *Advanced Computational Intelligence: An International Journal (ACII)*, 3 (1): 37–47.

- Ware, M., and Mabe, M. (2015), *The STM Report: An Overview of Scientific and Scholarly Journal Publishing*, The Hague, Netherlands: International Association of Scientific, Technical and Medical Publishers.
- Werder, K.P., Nothhaft, H., Verčič, D., and Zerfass, A. (2018), "Strategic Communication as an Emerging Interdisciplinary Paradigm", *International Journal of Strategic Communication*, 12 (4): 333–351.