Lingue e Linguaggi
Lingue Linguaggi 59 (2023), 333-349
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v59p333

AN EXPERIMENTAL CRITICAL MULTIMODAL DISCOURSE STUDY TO THE AI-DRIVEN SENTIMENT ANALYSIS OF ONLINE CRISIS COMMUNICATION

CARMEN SERENA SANTONOCITO, CHIARA POLLI[1]
UNIVERSITÀ DI MESSINA

Abstract – In response to the challenges of crisis management and communication in business digital scenarios (Umar *et al.* 2022; Catenaccio 2021), this case study presents an example of the evolution and implications of online crisis communication and discursive practices that combine human input and AI for the management of crisis events and trust repair. Reliability of communication in web-based business scenarios cannot be easily achieved because of the huge amount of data involved. In addition, reputation management and trust are constantly put under threat by misinformation and misunderstanding (Garzone, Giordano 2018). With a view to countering these risks, companies are increasingly outsourcing digital marketing services (Palttala *et al.* 2012) based on AI methods to analyse consumers' online needs and behaviour (Schwaiger *et al.* 2021), even though research has recently raised some concerns about over-reliance on AI-based tools (Tam, Kim 2019). We intend to show how multimodal critical discourse studies (Djonov, Zhao 2014) can help identify and understand the potentials and limits of social media listening tools that are designed for crisis prediction and management (van Zoonen, van der Meer 2015). To do so, we present a case study that we engaged with during our research traineeship at Digital Trails, a B2B company dealing with online visibility, digital marketing and reputation analysis. In reporting the outcomes of online reputation analysis to gauge possible damage after a crisis event, we present a comparison between AI-driven sentiment analysis (conducted via Meltwater) and the human-based revision and fine-tuning of AI-driven sentiment analysis. Our aim is to discuss the potentials and the criticalities of AI-driven sentiment detection. Among the latter, we highlight the unrecognizability of languages other than English, and flaws in interpreting pragmatic aspects, as well as multimodal digital artefacts. Consistent with our findings, we argue that AI models based on a unimodal and decontextualized architecture still require human validation. We conclude by indicating research directions for the detection of sentiment polarity, which include higher collaboration between IT developers and multimodal discourse analysts so that multimodally-informed models can assist crisis communication and management more efficiently.

Keywords: online crisis communication; social media listening tools; critical multimodal studies; AI-driven sentiment analysis; multimodality and AI.

1. Background and current issues in sentiment analysis for crisis communication

The explosive growth of digital participatory platforms (Androutsopoulos 2013) and post-truth theories (Lewandowsky *et al.* 2017) has resulted in significative changes and in the harnessing of linguistic and broader semiotic resources (Diers-Lawson 2019) in online crisis management communication (OCMC, henceforth). In light of the current communicative shift in OCMC, companies need to monitor what potential customers think, which semiotic

[1] Although the authors cooperatively worked at the case study as a research team and discussed and conceived of the paper together, C. Serena Santonocito authored Sections 1, 2 and 5; Chiara Polli authored Sections 3, 4 and 6.

resources they use, and how they spontaneously express their thoughts, tastes and representations (Kotras 2020). Digital users' opinions and sentiments on specific products and services have become "the driving force in the bleeding edge of crisis communication" (Coombs 2014, p. 1) for their unpredictability and amplified resonance. In particular, user-generated contents (UGC henceforth) may stir uncontrolled public reception, such as the threat of viral misinformation and misunderstanding (Garzone, Giordano 2018), thus demanding proactive communicative responses (van Noort, Willemsen 2012).

In the face of the vast amount of social media conversations, user-generated blog posts, responses to brand-generated contents and online reviews, the use of AI-powered technologies to cope with extensive data provides a competitive advantage for companies (Schwaiger *et al.* 2021). As Morgan and Wilk (2021) explain, response to a potentially damaging event or general criticism is a key component of consumers' loyalty and brand equity. From a discourse-oriented perspective, the centrality of ethical concerns in business communication "suggests a move away from post-crisis communication mode to a business-as-usual, CSR-informed one" (Garzone, Catenaccio 2021, p. 140), whereby OCMC becomes part of established business discursive practices subsumed under the umbrella term of corporate social responsibility (CSR henceforth). In this light, gauging digital users' sentiments and opinions is a major concern for businesses to avoid and/or respond to potential crises to protect brand reputation and overall CSR. Technologies such as social media listening tools (SML henceforth) can be considered proprietary assets used by companies to monitor and analyse public online discourse on a variety of platforms (i.e., blogs, forums and social media) in real-time, allowing for automated sentiment analysis for crisis prevention and management (Perakakis *et al.* 2019). In the best case scenario, UGC monitoring can help identify emerging issues and potential crises before they escalate (Coombs, Holladay 2012), whereas, in the event of a crisis, public sentiment and perception from key stakeholders and influencers can help assess the crisis impact on the brand reputation and implement OCMC strategies accordingly (Liu *et al.* 2011).

From an IT viewpoint, automated sentiment detection incorporates AI models into SML tools by harnessing in-built techniques, such as Natural Language Processing (NLP henceforth), text mining and information retrieval (Al-Ghamdi 2021). These techniques allow, *inter alia*, to process large-scale data to perform AI-driven sentiment analyses of such unstructured data, which usually come from different channels and diverse semantic levels (Garbade 2018).

Studies of AI-powered SML tools are well attested in disciplines such as business communication and computer science. In the former domain, Buzoianu and Bîră (2021) incorporate predictive models of crisis analysis to the use of SML to detect visibility, media coverage and sentiment of digital users. In light of the current flaws in automated sentiment analysis, they compare the latter with manual sentiment detecting. In a similar vein, Hayes *et al.* warn against the perils of "relying on automatically generated results from black-box SML platforms" (2021, p. 89). In the latter domain, Schwaiger *et al.* (2021) recognize challenges related to the focus on only one analytical approach, and to the handling of multifarious semantic levels in verbal data. In a similar vein, Bukar *et al.* (2022) highlight that current Machine Learning (ML henceforth) techniques lack satisfactory linguistic support for data monitoring and classification, as well as poor domain-specific datasets. Although these studies recognize the key role of SML platforms in dealing with OCMC, weaknesses in classification, processing, and interpreting of big data, including multimodal components, still need to be satisfactorily addressed.

This paper is structured as follows: Section 2 overviews research literature, presents current issues related to AI-driven sentiment analysis and introduces our RQs. Section 3 illustrates methods and data, while Section 4 presents the analysis. Section 5 discusses the results and addresses our RQs. Section 6 provides some concluding remarks and indicates possible future lines of research.

2. Theoretical framework and research questions

In this paper, we wish to address the challenges of incorporating multimodal data in ML for OCMC from a Critical Multimodal Studies (CMS henceforth) standpoint (Djonov, Zhao 2014). In our approach, multimodal data are not simple channels for communication but are instead conceptualized as semiotic resources for meaning making within the tradition of social semiotic studies (Kress 2010).

Since we collected multimodal data in digital textualities, we embrace the notion that current digital semiotic resources (including, but not limited to, verbal texts, typographic devices, social distance in web-based interactions) combine in unprecedented ways to create multimodal meaning-making (Sindoni 2013), including the formation of sentiment.

In light of the foregoing observations, the role of digital textualities, as well as the sentiment therein conveyed, share some commonalities with the social constructionist notion of *discourse* – i.e., a constitutive process that shapes our knowledge about particular topics and events by making available certain linguistic and extra-linguistic representations and procedures, while excluding others (Foucault 1965) – and with the CMS tradition. CMS-oriented studies (Kress, van Leeuwen 2020; Zhao *et al.* 2018), inspired by the Foucauldian notion of discourse, have come to conceive the latter as multimodal for it is going beyond the verbal mode. As a matter of fact, according to CMS discourse encompasses all semiotic resources that help construct worldviews (Djonov, Zhao 2014) and create polarized sentiment towards certain events in specific social contexts.

Consistent with CMS tenets, terms such as *textuality* and *discourse* are to be intended in this study as the multimodal orchestration of different communicative resources to create meanings, including sentiment, in response to socio-historical interests and power configurations (Djonov, Zhao 2018). Furthermore, the CMS perspective adopted will allow us to address the major concern that AI models used to train SML algorithms are mostly unimodal (Das, Singh 2023; Gandhi *et al.* 2023). This implies that algorithms which detect sentiment consist of language-alone datasets with little (and problematic, as we will show in this paper) or no training apt to analyse and interpret other semiotic resources, such as images or aural resources. As a consequence, AI models are likely to fail to recognize (or misinterpret) the intended meanings and sentiment resulting from a combination of resources and not from one single resource (Tam, Kim 2019; Davidson *et al.* 2017; Duarte *et al.* 2017; Majumder *et al.* 2018).

Against the backdrop of an increasing interest in having access to opinions, beliefs, emotions and perceptions of the public (Kotras 2020), the growing research devoted to sentiment analysis (Di Cristofaro 2023) is not surprising. Sentiment analysis refers to a domain relevant to computer science and computational linguistics, which focuses on the polarity of sentiment, underlying "positive or negative feeling implied by opinion" (Liu 2017, p. 12) expressed by means of verbal language. If we consider AI-driven sentiment analysis, current research highlights an additional broader issue, defined "context blindness" (Dias Oliva *et al.* 2021, p. 3). The latter is considered as a further limitation of AI models used for online sentiment recognition in SML platforms because current AI sentiment detection uses unimodal verbal protocols that process an incomplete message. As a result, the AI-driven sentiment does not account for the combination of other resources (such as images) for meaning making in digital texts (Moschini, Sindoni 2022). The implication is that semiotic resources other than language are disregarded by AI-powered SML tools, therefore ultimately producing an incomplete or misinterpreted message to the point of misunderstanding the inherent sentiment.

To address these challenges, this paper draws on CMS to discuss fieldwork on a case-based study of OCMC, scrutinized by means of the SML platform Meltwater. We used Meltwater to address this experimental case study during a research training programme at the international digital marketing agency Digital Trails (DT henceforth).[2] Our case study aims to contribute to CMS-oriented research for the development of multimodally-informed AI protocols.

By comparing the AI-driven and the manual sentiment analyses conducted after a crisis event, we address the following research questions (RQs):

- RQ1: What are the limits and potentialities of AI-powered sentiment analysis in SML tools such as Meltwater?
- RQ2: Does the SML tool successfully assess sentiment of multimodal combinations?

3. Materials and Methods

In the wake of the increasing interest in empirical analyses within the field of CMS (Bateman *et al*. 2017; Jewitt 2014), our reflections are based on empirical data that we collected during a 6-month field-work experience (07/03-06/09/2022) at DT. The general aim of the experience was to develop a study on AI based on *real-world* data and in synergy with external partners, whose expertise and practical knowledge of AI applications have certainly broadened, in our experience, the scope of our academic research.

Our partner is a SME, DT, which operates worldwide with a varied range of clients (from education to online retailers, business-to-business and business-to-consumer services) that aim to build, manage, and restore their e-notoriety. We conducted a field-work experience in collaboration with DT's core team of marketing experts and assisted them in tasks related to online visibility and online reputation management, such as digital marketing and PR, pay-per-click campaigns, online brand mention monitoring, and Search Engine Result Page Analysis (SERP henceforth). In particular, we were involved in OCMC tasks aimed at assessing reputational risks and damage, counteracting potential harm caused by negative reputational events, and improving the clients' search profiles.

The case-study discussed in this paper was part of an online reputation audit task carried out between April and June 2022 and involved an international brand providing fermentation products, here anonymized. The crisis event, which led the company to contact DT, was the sale of a faulty batch of bottles of wine, which caused several fermentation problems for customers (e.g., gushing bottles). The client required an overview of the search landscape in its three main markets (i.e., UK, US, France) based on a set of agreed brand, product, and generic keywords. In addition, DT was asked to identify the online mentions of the faulty batch incident, which may have resulted in negative feedback from potential customers, and to provide insights into the sentiment of relevant online conversations about the company.

DT's intervention was split into a four-step process. The first step consisted of selecting the most efficient keywords that would generate comprehensive results. DT's

[2] Our activity was supported by the Italian National Ministry of University "Development of linguistic and semiotic protocols for Artificial Intelligence models informed by multimodality" within PON "Programma Operativo Nazionale Ricerca e Innovazione" 2014-2020, Azione IV, 4, "Dottorati e contratti di ricerca su tematiche dell'innovazione", 01/01/2022 – 31/12/2024. CCI2014IT16M2OP005.

experts opted for the brand's and the faulty products' names, and some generic but relevant search terms (e.g., *wine making yeast* and *levure de vinification*) in English and French (i.e., target markets' official languages). The second step involved the use of the SML tool Meltwater to analyse the online landscape for the three target markets across a wide variety of potential sources, such as product review websites, blogs, news sources (e.g., New York Times, BBC, CNN, local newspapers, online journals, and content posted on television networks' websites), social media platforms (Facebook, Twitter, Reddit, TikTok, Instagram, Pinterest), video live streaming and sharing platforms (YouTube and Twitch).

Meltwater is an AI-powered social media monitoring and listening tool listed among the business analytics instruments most frequently used by contemporary organizations (Ziora 2016). Meltwater allows users to refine audit searches by using Boolean operators and the filtering bar, which discriminates results according to source types (e.g., blogs, Twitter, Reddit *etc.*), language (language detection for 242 languages is supported and full sentiment analysis for 28 languages is provided),[3] location, other keywords, sentiment, author (e.g., lists of influential users of Twitter, Reddit *etc.* can be created), and other customised categories. By using Meltwater, DT experts extracted publicly available data for the types of keywords analysed. Boolean search operators (particularly, *and*, *or*, and *not*) helped process relevant mentions based on different keywords. Generic keywords were used in correlation with the brand and product names to further expand the search. Data were filtered according to location to collect mentions in the client's three core markets (UK, US, and France). No language filter was used so as to include potential mentions of the client's brand and product in languages other than those which are official in the target countries. As a third step, the results gathered were categorized according to their sentiment, both via Meltwater and by conducting a manual annotation. This is the step on which the study discussed in this paper is based. The fourth and last step of DT's task regarded a topline SERP analysis, that is the analysis of the first two pages of Google search for the three main target markets, with the aim of detecting potential negative mentions in prominent positions. We manually conducted the sentiment analysis of SERP data, but the results are not relevant for the research purposes of this paper, thus they will not be discussed here.

Once completed this 4-step assessment of the client's reputation following the crisis event, the DT's core team met with the company's spokesperson and discussed the results as well as further potential actions to boost the client's online presence and e-notoriety.

As mentioned above, we were mainly involved in the third step of DT's crisis management task, i.e., the sentiment analysis phase, which we used as a testbed to assess the accuracy of AI-driven sentiment analysis in comparison to a manual revision and fine-tuning. As mentioned above, the former was conducted by using Meltwater. This tool can assess sentiment via NLP, that is, based on the computational process of automatically analysing human language by training the algorithm to recognize the polarity of texts and categorize them according to three sentiment labels, namely, 'positive', 'neutral' and 'negative'. The result is labelled as 'not-rated' whenever the system fails to recognise the type of online contents.

The overall sentiment of a text is determined by a polarity score. As per their guidelines, "if something is both positive and negative, but more positive, the polarity score would rank it as positive" (help.meltwater.com). In addition to language, Meltwater is also programmed to detect emoticons and emojis. Owing to its proprietary nature, further details on Meltwater's data processing and algorithms were not shared by its spokespersons collaborating with DT.

3 https://help.meltwater.com/en/articles/4064558-how-is-sentiment-assigned.

In this analysis, Meltwater retrieved and automatically categorized 4,804 items according to their sentiment, covering the 12 months which followed the crisis events. Subsequently, we manually refined results by eliminating double entries, with a total of 2,567 unique items that constitute the dataset used for the analysis.

Entries referred to a wide variety of digital text types, including recipe and personal blog posts, brewing industry-related websites (magazines, journals, and suppliers' webpages), buying guides, review websites, Twitter posts, YouTube videos, 6Parks.news and Player.fm podcasts, as well as a selection of pages from the client's own website.

We then manually reviewed and tagged for sentiment analysis all items. In addition to the labels *positive*, *negative*, and *neutral*, we were asked to use the categories *not accessible* (in case of broken links) and *irrelevant* (for spam or mentions related to different brands), since Meltwater was unable to discern automatically these types of information.

To improve inter-annotator reliability, we analysed all items separately. A third annotator from DT also tagged the results independently. The output of the three different analyses were subsequently compared and, in cases of disagreement, the senior annotator attributed the sentiment according to the company's in-house decision-making strategies.

The final dataset was therefore compared with the output of AI-driven sentiment analysis performed via Meltwater and all items were qualitatively re-examined by performing a content analysis to explore and categorise potentially recurring criticalities.

4. Results and Findings

AI-driven sentiment analysis	
Rating	Items
Positive	161
Neutral	2,358
Negative	12
Not Rated	36

Table 1
AI-driven sentiment analysis.

Human-based revision	
Rating	Items
Positive	179
Neutral	2,069
Negative	4
Irrelevant	59
Not Accessible	256

Table 2
Human-based revision.

As shown in Table 1, Meltwater produced positive results in 161 cases, neutral in 2,358 cases, negative in 12 cases, while 36 items were not rated. By contrast, human-based revision produced different results: 179 items were labelled positive, 2,069 neutral, 4 negative, 59 irrelevant, 256 not accessible. The discrepancy between the results is even greater, since there was no full agreement between what was considered positive, neutral and negative by the software and by us researchers. For instance, the majority of positive mentions (67.5%) manually detected were rated differently by Meltwater: in particular, positive mentions were considered neutral by AI in 64.8% cases, negative in 1.7% cases, and not rated in 1%. As for negative results, which were the most relevant for the analysis of the client's crisis event, the AI failed to recognize all 4 negative mentions detected by human annotators: in particular, negative results were labelled positive by AI in 25% cases and neutral in 75% cases. This also indicates that no negative result detected by Meltwater was considered negative in our manual revision. Neutral results were less problematic, though 5.7% of the items we considered neutral were either considered positive (4.2%) or not rated (1.5%) by AI. In the majority of cases, the neutral results were not particularly challenging and referred to blog recipes and buying guides, in which the keywords were mentioned among the lists of ingredients for a given recipe, special offers and deals, and product compositions.

We subsequently explored the dataset qualitatively, by focusing on the entries where we found a mismatch between the human revision *vs.* the AI-driven analyses, so as to highlight which types of challenges Meltwater may encounter. We classified the main criticalities in three broad categories, which are presented in the following Sub-sections:

4.1. Contents in languages other than English;

4.2. Contents whose interpretation requires specific contextual and pragmatic skills;

4.3. Multimodal contents.

4.1. Contents in languages other than English

As mentioned in Section 3, Meltwater enables to refine data search based on location and language. In this case, DT was interested in data from three target markets (UK, US, and France), but did not select entries based on language so as to include potential mentions than were not in the official languages of the target countries.
As shown in Table 3, English was found to be the predominant language in the dataset for search results related to the clients' brand and its products (95.17%).

Language breakdown	
Language	Percentage
English	95.17%
Spanish	1.05%
French	0.66%
Portuguese	0.47%
Moldavian	0.43%
German	0.39%
Russian	0.35%
Swedish	0.19%
Italian	0.16%
Chinese	0.12%
Polish	0.12%
Czech	0.04%
Finnish	0.04%
Gujarati	0.04%
Turkish	0.04%
Unknown	0.74%

Table 3
Language breakdown.

The remaining entries included texts in other 14 languages (4.09% of the total) and not rated results (0.74%). The latter result refers to YouTube video contents which, as indicated in Sub-section 4.c, could not be rated by Meltwater. By contrast, Meltwater was able to correctly identify all 14 languages and correctly labelled 56.8%[4] of the results expressed in languages different from English (mainly neutral blog recipes and buying guides). However, the fact that in the remaining 43.2% cases there was a mismatch between our revision and the AI-driven analysis confirms that sentiment assessment in these languages may be tricky for AI.

[4] Non-accessible results were excluded from the count.

The manual and qualitative analysis of these results confirmed this hypothesis. Entries in French represent a significant example in this respect: for instance, 50% (i.e., 2/4) of the negative mentions detected in our analysis included the articles from two online consumer advice magazines in French, which were mislabelled as neutral by AI. This inaccuracy was particularly problematic, because France was one of the clients' target markets and negative mentions are clearly more relevant than neutral ones to evaluate reputational damage. By looking at keywords and cues of polarity, we found that the articles explicitly referred to the crisis event under scrutiny and the risks for buyers it had represented (e.g., *risque d'explosion*, En. Translation: risk of explosion; *risques encourus par le consommateur*, En. translation: risk incurred by the consumer). Since the magazines aimed to inform clients about the crisis event, the articles had to be explicit and unambiguous. In this respect, the fact that Meltwater mis-labelled them despite the lack of tricky items or pragmatic ambiguities seems to suggest that it may be less precise in the sentiment analysis of texts in French. This seems also confirmed by the fact that 66.7% of the positive mentions in French (promotional articles in online brewing industry magazines) were labelled as neutral regardless of positive polarity keywords such as *très appréciées* (En. Translation: much appreciated) and *popularité croissante* (En. Translation: growing popularity) referring to the client's brand and products.

4.2. Contents whose interpretation requires specific contextual and pragmatic skills

Another area of concern regards entries, such as Twitter posts and personal blog articles, in which users tended to express personal opinions by adopting a more informal register, often combining positive and negative statements, and using sarcasm, irony, slang terms, idiomatic and unusual expressions. The dataset abounded with these text types, which were unambiguous for us to interpret, but certainly proved challenging for machines as interpretation required interlaced contextual and pragmatic skills.
These issues account for the wrong labelling of the remaining 50% of negative mentions (2/4), which are both tweets discussing the users' personal experiences with the client's products.

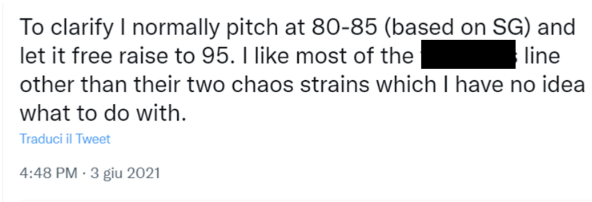

Figure 1
Negative sentiment rated Neutral by AI.

In the example included in Figure 1, the user described how they appreciate the client's products "other than their two chaos strains" claiming that, after the crisis event, they "have no idea what to do with" the client's products. Despite the use of the positive polarity keyword *like*, we considered the tweet a negative mention, because it referred to the crisis event and to the unusable products. However, Meltwater labelled the text as neutral, thus disregarding all indirect negative allusions.

In the second case, the entry was part of an exchange of tweets between two users discussing the client's product in comparison to those of a competitor. Since Meltwater selected and analysed only the tweet in which the client's brand is mentioned (Figure 2), we limited our analysis to the same post.

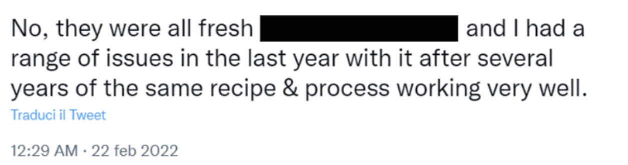

Figure 2
Negative sentiment rated Positive by AI.

In the text in Figure 2, the user refers to "a range of issues" connected to a client's product, despite it was said to be "working very well" in the past. Despite the presence of potentially positive keywords ("fresh", "very well"), we interpreted the tweet as a negative mention of the product as it is implied that the user was no longer satisfied. However, the interpretation of this text proved complex for Meltwater, which categorized it as positive in light of the positive polarity keywords and the lack of cause-and-effect logical skills needed to understand that these keywords referred to a past, pre-crisis event situation.

Figure 3
Positive sentiment rated Neutral by AI.

Furthermore, in several cases, Meltwater struggled with items which required the interpretation of irony and sarcasm frequently conveyed by using slang terms and idiomatic expressions. For example, a wholesaler's tweet (Figure 3) included the sentence "these sales are too hot to handle" together with the client's brand and products name and a link to the wholesale online shop. The idiomatic expression *too hot to handle* was interpreted by Meltwater as a reference to something "so dangerous, difficult, or extreme that people do not want to be involved with them" (Collins Dictionary), thus leading to a negative rating. Conversely, in this case, the tweet is actually conveying a positive sentiment, as the expression is rather used ironically to promote an extremely appealing sale.

Likewise, in another tweet (Figure 4), the client's product was said to be "KICKING" (in capital letters) by a user who displayed enthusiasm for the fermentation process going on. Meltwater failed to detect the meaning of the message and interpreted it as neutral rather than positive. The capitalization of "KICKING" was part of a visual typographical strategy to enhance the salience of the term, thus altering the overall meaning-making (Kress, van Leeuwen 2020). In addition, the tweet included a short video-clip of the ongoing fermentation process. However, these features cannot be detected by AI-driven unimodal classifiers, thus calling for the incorporation of multimodal data for a more fine-grained comprehension of sentiment.

Figure 4
Positive sentiment rated Neutral by AI.

4.3. Multimodal Contents

As anticipated in Section 2, given the inherently multimodal nature of digital communication, sentiment is frequently conveyed by the combination of multiple semiotic resources, which may or may not include verbal language. For this reason, we accounted for verbal and non-verbal resources (as well as their interplay) in the manual annotation process. The qualitative analysis of the differences between our results and those of Meltwater revealed the poor performance of AI in comparison to manual classification when dealing with texts whose sentiment is conveyed by a combination and/or interplay of several resources.

For example, in several entries in the dataset, sentiment was conveyed by graphic elements, such as star-rating systems, emoticons, and emojis. Star-rating systems were usually added next to the client's products in buying-guide and e-commerce retailer pages. In all the cases examined, the products' rating was high (4+ stars) and, therefore, we labelled the mention as positive. In most cases, the presence of the visual mode proved fundamental to determine the sentiment, since the text alone often presented no polar word and thus would be classified as neutral.

In AI-driven sentiment analysis, results were influenced by the fact that the visual mode was disregarded: as a consequence, all texts in which the sentiment was conveyed by the star-rating system were classified as neutral or even negative.

Even though Meltwater is trained for emoji detection, the qualitative examination of sentiment analysis results showed a poor emojis categorization. As shown in Figure 5, several entries, in which one or more positive emojis were included, were actually mis-labelled by the system.

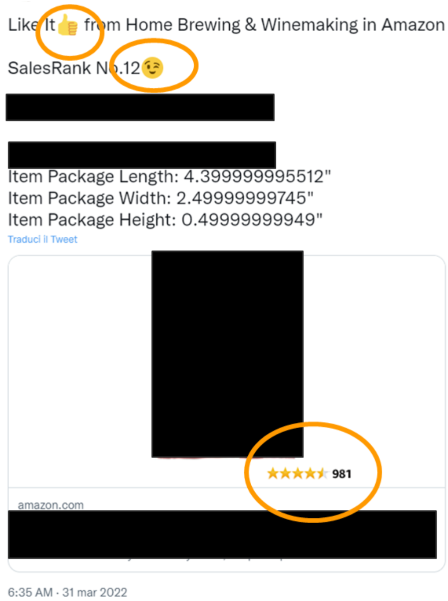

Figure 5
Positive Sentiment Rated Negative by AI.

In particular, in the Twitter post in Figure 5 the positive sentiment is conveyed by the 'like' symbol and the winking face emojis, as well as by the star-rating system. Meltwater, however, failed to recognize these elements and classified the mention as negative.

Further issues are linked to contents (i.e., videos and vidcasts) from video-sharing platforms, particularly YouTube. In most cases, in both the videos and the captions, either the client's products were included in the ingredient lists of video recipes or the brand was mentioned as a sponsor and promoted by the content-creator(s) of vidcast interviews and chats. These mentions were labelled as 'neutral' and 'positive', respectively. By contrast, the AI failed to recognize this type of contents and all links to YouTube videos were always labelled as not rated, even though the brand's and the products' name were mentioned in the captions.

5. Discussion

Our experience at DT provided fruitful insights into the significance and the challenges of harnessing AI protocols for OCMC. Indeed, we had the opportunity to understand how the study of AI-powered SML tools extends its relevance beyond computer science to incorporate discursive and multimodal foci as well. In trying to answer our initial RQs, we bear in mind this cross-contamination of disciplines that intervene for reciprocal enhancement of the discussion, despite their inherent differences.

While acknowledging some IT aspects relevant to our discussion, we focus on multimodal and discourse-oriented elements when addressing RQ1. Tables 1 and 2 show that the comparison of human revision *vs.* AI-driven sentiment detection delivers a fairly approximate indication on the distribution of sentiment. Indeed, in mere terms of statistical significance, the figures of positive, negative and neutral mentions extracted by the AI-driven sentiment analysis (Table 1) do not significantly differ from the overall manual categorization of human-based revision (Table 2). If we consider this aspect on its own, the merit of AI-powered SML tools, such as Meltwater, rests on the production of a general quantitative overview on sentiment classification over large-scale unstructured data. ML retrieval, indexing and assessing of massive amounts of data in real-time and from different web sources contributes to minimizing human workforce and reducing operational costs.

Another benefit consists in the possibility to refine searches through the various customization features (e.g., filter the location of the market, the social network; isolate a source of information, the kind of communication – B2B, B2C *etc.*) and through the Boolean operators, so as to facilitate processes of OCMC and decision making. Agreeing with Perakakis *et al.* (2019), our experience confirmed that AI-powered SML technologies ease the automation of various marketing tasks, large-scale monitoring of sentiment, and human efforts reduction.

Aware of the fact that SML technologies are market-driven products, we recognize that these tools are nonetheless contributing to the field of digital social sciences. From our field work experience, we understood the extent to which SML tools are shaping companies' departments and expert knowledge on brand reputation and risk assessment. Technologies like Meltwater are also contributing to the process of hybridization of opinion measurement (Kotras 2020), currently interwoven among digital PR and various kinds of automated public opinion assessment. In some cases, SML tools are being incorporated into companies' internal resources; for example their potential is harnessed in marketing departments to support brand image management, digital PR, and customer service. Consistent with Bukar *et al.*(2022), AI-powered SML technologies can be considered as shaping corporate departmental organization and managerial decision-making for their centrality to sustainable OCMC.

As for their limits, examples in Section 4 show that, when dealing with sentiment detection, AI technologies tend to overlook a number of elements mainly related to linguistic and discourse-oriented aspects. In fact, even if the general distribution of the sentiment may indicate a trend that is consistent in both AI-driven and human-based analyses, our findings revealed considerable discrepancies between the two analyses when comparing the outcomes of positive, negative or neutral sentiment of each entry. These flaws cannot be fully explained due to the inaccessibility of the full architecture and operational functioning of SML tools; therefore, we cannot currently track down the rules (e.g., algorithms) that discriminate among positive, negative and neutral labelling in the SML protocols. Echoing Schwaiger *et al.* (2021) and Hayes *et al.* (2021), we draw on the metaphor of SML technologies as black boxes. Being patent-based, these tools obscure how data are collected and processed.

Among the limitations mainly concerning the verbal mode, we focused on cases involving languages other than English, and cases requiring specific pragmatic skills and contextual knowledge. In the first instance, although Meltwater supports sentiment analysis for 28 languages, our results in Sub-section 4.a show that the SML tool correctly identified all the different languages but did not properly recognize the sentiment when languages other than English were involved. In the latter cases, human assessment was necessary for the correct identification of sentiment. This flaw in AI-driven sentiment detection was fairly remarkable for textualities in French, a language widely spoken in one of the client's target markets. Indeed, Meltwater language models were not able to detect negative sentiment, nor explicit mention of the crisis event in the French language. Because of recent discussions on the linguistic imperialism of English as the dominant language in which sentiment analysis has been originally developed (Joshi *et al.* 2017; Kiritchenko *et al.* 2016), it might be the case that in our data the parallel multi-lingual mapping of words reproduces the first computational models developed for the English language. Despite the relative simple textual contents in French, these criticalities have nonetheless emerged (See Sub-section 4.a).

In the second instance, specific human-based interpretations were needed for correct sentiment detection, which requires the simultaneous combination of lexical, syntactic, semantic and pragmatic dimensions. Among the examples presented in Sub-section 4.b, we distinguish between limitations caused by low display of sentiment (Kotras 2020) and by non-linear pragmatic and semantic levels. As for the low intensity of sentiment, examples in Figures 1 and 2 show how the SML tool failed to detect the correct sentiment where non-extremely polarized words appeared, or where negative and positive words were equally present at the sentence level. As for pragmatic and semantic non-linearity, Figures 3 and 4 highlight that AI models seem to lack training to recognize slang and friendly-tuned language, in addition to semio-pragmatic competence (Odin 2022) to understand the semantic marking of typographic devices, such as capitalization (Chan, Fyshe 2018). In fact, the AI rated neutral sentiment in Figures 3 and 4, while our qualitative classification marked sentiment in both figures as positive.

Another limit concerns the algorithm inability to recognize inaccessible web addresses, spams and homonymous mentions related to other events. In our experience, the SML tool mainly labelled them as neutral, thus jeopardizing the general sentiment distribution.

As for RQ2, we showed that the SML tool struggled to detect correct sentiment when different semiotic resources combined to create meaning. As illustrated in Figure 5, the sentiment conveyed by the interplay of the verbal mode with other visual resources, such as emojis, pictures, and icons was not correctly understood by the AI. Figure 5 indicates that, in some cases, the output of AI-driven sentiment analysis was misleading because it assessed as negative the overtly positive use of icons, as well as combinations of text and emojis.

The same can be said about the incorporation of the aural mode; indeed, the sentiment expressed in podcasts and other oral resources was misinterpreted by the AI. Concerning video-sharing platforms such as YouTube, although the latter is quoted among one of the sources for sentiment detection in Meltwater, its multimodal content (e.g., ensemble of aural, visual and verbal modes) was not recognised: all YouTube contents were labelled as *not rated*, despite the overtly positive lexis used in captions and hashtags. In such multimodal data, the algorithm did not even recognize as positive verbal mentions of the product’s brand. Even if mentions might be considered as neutral sentiment, these actually display low positive sentiment. In some cases, these occurrences are represented in the verbal mode, whereas in other cases these are reproduced in the visual or in the aural mode. However, the fact that the algorithm ignores such low displays of positive sentiment, be they in the verbal or in other modes, obscures positive mentions. As a result, the algorithmic misinterpretation of sentiment moves to the background good marketing practices. Furthermore, our findings indicate that the AI-driven sentiment detection did not account for the hyper-textuality of social networks (i.e., Twitter, YouTube). As a matter of fact, in none of the examples shown in Section 4 the SML tool considered ‘likes’, visualisations, and reposts for sentiment assessment, thus ignoring dynamics related to online engagement and algorithmic visibility of brand mentions.

Since in all the examples of Section 4 the joint consideration of pragmatic elements and multimodal aggregations was essential for correct sentiment detection, the AI-powered sentiment analysis of our dataset was compared and subsequently reviewed following human revision and validation. Should such revision not have been carried out, we would have compromised one of the main tasks, i.e., online reputation audit by not signalling negative mentions, nor enhancing the best digital marketing practices.

In line with the CMS perspective of this study, we consider challenging aspects related to the artificiality and automation of SML tools as criticalities that still need to be adequately addressed. These criticalities are mirrored in the inaccessibility of algorithmic architecture and in the predefined setup of Boolean operators, which tend to produce a standardized set of questions and answers. We contend that such blind automation can alter context-dependent evaluations on specific events and brands, since the visibility of online opinions hardly corresponds to the beliefs of average population (Kotras 2020). Following Fairclough (1992) on the technologization of discourse, and Djonov and van Leeuwen (2012) on semiotic software products, we also wish to encourage reflection on the risk of over-relying on automatization and top-down imposition of AI-driven sentiment detection. If, on the one hand, automated computations are fast and cost-effective, on the other hand, reliance on unimodal protocols impinges on automated processing of sentiment. Inasmuch CMS call for the scrutiny of automated software products in light of broader social and cultural contexts (Djonov, Zhao 2018), so we argue that the models and the output of AI-driven sentiment analysis should be re-evaluated in relation to human validation and contextual and pragmatic knowledge of the world.

Following our case study, and recalling the impact of digital sentiment on business communication and on human behaviour (Liu 2012), we acknowledge the centrality of AI-powered SML tools from a theoretical perspective – e.g., IT and social sciences – and from a practical one – e.g., more efficient OCMC. Subsequently, we argue that the application of AI-driven sentiment analysis in various business, governmental and social domains consider the necessity to both cross disciplinary boundaries, and to include multimodal discourse analysts to refine AI sentiment assessment. In particular, given the fruitful potential of our experience at DT, we suggest an integrated approach between AI developers and multimodal discourse analysis so that in-built data collection, labelling and training of SML tools can be informed by CMS expertise and validation for enhanced OCMC.

6. Conclusions

In this paper, we addressed the task of sentiment analysis by discussing an OCMC case-study based on our direct field-work experience. We decided to conduct a parallel AI-driven *vs.* manual analysis of a newly created dataset gathered online by using the SML tool Meltwater, which was also used to carry out the subsequent sentiment analysis phase. The dataset consisted of positive, negative, and neutral mentions of the brand and products involved in a crisis event. It included texts in 15 languages from three geographical areas (UK, USA, and France). In this case-study, we were able to delve into the CMS-informed study of the challenges SML tools encounter when dealing with the analysis of digital communication, by qualitatively examining several types of textualities and different forms of integration of visual, audio, and textual features. Although we recognize the invaluable contribution of AI tools in collecting and classifying huge amount of data, considering our results and the CMS perspective, we ultimately maintain that the performance of AI-driven tools (in this case, Meltwater) can still improve significantly if two types of primacies are rediscussed: first, the prioritization of English (even though with the caveat that most available big data are in English); second, the prioritization, of (written) language among the different modalities of communication. As for the former, our results indicate that AI still suffers from a bias caused by the linguistic imperialism of English, to the detriment of other languages (as well as English slang forms). As for the latter, we showed how the proper classification of polarity often depends on the interpretation of multimodal aggregations and, thus, AI based on unimodal – in this case, language only – classifications, with a null or limited integration of other modalities, are deem to yield unreliable results. In addition to these areas of concern, a final criticality revealed by our analysis is the lack of pragmatic understanding of multimodal data. What is straightforwardly understood by humans can be extremely challenging for machines to detect. We acknowledge that this study is based on an experimental research on a single SML tool and on a limited case study. While additional discourse-oriented research is needed to further disentangle the challenges of AI-driven sentiment analysis, we believe our experience has showed the potentialities of this research direction and can encourage novel investigation with datasets covering different domains and analysed by using different SML tools.

Bionotes: Carmen Serena Santonocito is research assistant in English Linguistics and Translation at the University of Messina, Italy. She holds a PhD in European Languages and Specialized Terminology from the University of Naples "Parthenope". Her current research on the criticalities of multimodal protocols used in artificial intelligence is funded by the PON project "Development of linguistic and semiotic protocols for Artificial Intelligence models informed by multimodality". Her research interests include Gender Studies, Corpus-assisted Critical Discourse Studies, Critical Multimodal Studies and Terminology. She has published in international double-blind peer reviewed journals such as *CADAAD*, *Anglistica AION*, *MediAzioni*.
Chiara Polli is Research Fellow of English Linguistics and Translation at the University of Messina, Italy. She was awarded her teaching and research position within a three-year competitive national-based project on "Development of Linguistic and Semiotic Protocols for Artificial Intelligence Models Informed by Multimodality", which combines linguistics and sociosemiotics to computer science approaches to address the challenges of AI-based automatic detection of harmful text-image clusters. Her research interests include critical discourse studies, multimodality, semiotics, hate speech, digital communication, and translation. Her works have been published in international journals such as *TTR: Traduction, Terminologie, Rédaction* and *Punctum. International Journal of Semiotics.*

Authors' addresses: carmenserena.santonocito@unime.it; chiara.polli@unime.it

Acknowledgements: We wish to thank Digital Trails (https://www.digitaltrails.co.uk/) for having us in a traineeship programme within the PON "Programma Operativo Nazionale Ricerca e Innovazione" 2014-2020. We would like to thank in particular Dr. Bogdan Marinescu, who has generously shared his experience and expertise with us. We are aware that the combination of our background as linguists and multimodal discourse analysts and the technical know-how we gained at Digital Trails has been fundamental to design and carry out field work of this kind. We are grateful for this opportunity that has greatly enhanced our experience especially in the field of AI-driven models for the analysis of language, as well as of other resources.

References

Al-Ghamdi L. M. 2021, *Towards adopting AI techniques for monitoring social media activities*, in "Sustainable Engineering and Innovation" 3 [1], pp. 15-22.

Androutsopoulos J. 2013, *Participatory culture and metalinguistic discourse: Performing and negotiating German dialects on YouTube*, in Tanner D. and Trester A. M. (eds.), *Discourse 2.0: Language and New Media*, Georgetown University Press, Washington, pp. 47-71.

Bateman J. A. *et al.* 2017, *Multimodality: Foundations, research and analysis: A problem-oriented introduction*, De Gruyter Mouton, Berlin/ Boston.

Bukar U. A. *et al.* 2022, *How Advanced Technological Approaches Are Reshaping Sustainable Social Media Crisis Management and Communication: A Systematic Review*, in "Sustainability" 14 [10], pp. 5854.

Buzoianu C. and Bîră M. 2021, *Using Social Media Listening in Crisis Communication and Management: New Methods and Practices for Looking into Crises*, in "Sustainability" 13, pp. 13015. https://doi.org/10.3390/su132313015

Chan S. and Fyshe V. 2018, *Social and Emotional Correlates of Capitalization on Twitter*, in *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media, Association for Computational Linguistics*, pp. 10-15. Doi:10.18653/v1/W18-1102

Collins Dictionary, https://www.collinsdictionary.com/dictionary/english/too-hot-to-handle (10.06.2023).

Coombs T. 2014, *State of Crisis Communication: Evidence and the Bleeding Edge*, in *Institute for public relations*. https://instituteforpr.org/state-crisis-communication-evidence-bleeding-edge/#:~:text=in%20crisis%20communication.,Social%20media%20is%20the%20driving%20force%20in%20the%20bleeding%20edge,is%20amplified%20by%20social%20media (10.06.2023).

Coombs T. and Holladay J. S. 2012, *The paracrisis: The challenges created by publicly managing crisis prevention*, in "Public Relations Review" 38, pp. 408-415.

Das R. and Singh T. D. 2023, *Multimodal Sentiment Analysis: A survey of methods, trends, and challenges*, in "ACM Computer Survey" 55 [13], pp. 1-38. https://doi.org/10.1145/3586075

Davidson T. *et al.* 2017, *Automated hate speech detection and the problem of offensive language*, in *Proceedings of the 11th international AAAI conference on web and social media*, https://arxiv.org/abs/1703.04009

Di Cristofaro M. 2023, *Corpus Approaches to Language in Social Media*, Routledge, London/ New York.

Dias Oliva T. *et al.* 2021, *Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online*, in "Sexuality & Culture" 25, pp. 700-732.

Diers-Lawson A. 2019, *Crisis Communication: Managing Stakeholder Relationships*, Routledge, Oxford.

Djonov E. and van Leeuwen T. 2012, *Normativity and software: A multimodal social semiotic approach*, in Norris S. (ed.), *Multimodality and practice: Investigating theory in practice through method*, Routledge, New York, pp. 119-137.

Djonov E. and Zhao S. 2018, *Social Semiotics. A Theorist and a Theory in Retrospect and Prospect*, in Zhao S. *et al.* (eds.), *Advancing Multimodal and Critical Discourse Studies. Interdisciplinary Research inspired by Theo van Leeuwen's Social Semiotics*, Routledge, New York, pp. 1-18.

Djonov E. and Zhao S. 2014, *From Multimodal to Critical Multimodal Studies through Popular Discourse*, in Djonov E. and Zhao S. (eds.), *Critical multimodal studies of popular discourse*, Routledge, London/ New York, pp. 1-16.

Duarte N. *et al.* 2017, *Mixed messages? The limits of automated social media content analysis*, in *Center for Democracy and Technology*, https://cdt.org/insight/mixed-messagesthelimits-of-automated-social-media-content-analysis. (14.06.2023).

Fairclough N. 1992, *Discourse and social change*, Polity Press, Cambridge.

Foucault M. 1965, *Madness and civilization*, Pantheon, New York.

Gandhi A., *et al.* 2023, *Multimodal Sentiment Analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions*, in "Information Fusion" 91, pp. 424-444. https://doi.org/10.1016/j.inffus.2022.09.025.

Garbade M. J. 2018, *A simple introduction to natural language processing*, in "Becominghuman. Artifical Intelligence magazine", https://becominghuman.ai/a-simple-introduction-to-natural-language-processing-ea66a1747b32 (20.06.2023).

Garzone G. E. and Catenaccio P. 2021, *Ethics in professional and corporate discourse. Linguistic perspectives*, Frank & Timme, Berlin.

Garzone G. E. and Giordano W. 2018, *Discourse, Communication and the Enterprise: Developments and Issues. An Introduction*, in Garzone G. E. and Giordano W. (eds.), *Discourse, Communication and the Enterprise: where business meets discourse*, Cambridge Scholars, Newcastle upon Tyne, pp. 1-17.

Hayes J. L. *et al.* 2021, *Can Social Media Listening Platforms' Artificial Intelligence Be Trusted? Examining the Accuracy of Crimson Hexagon's (Now Brandwatch Consumer Research's) AI-Driven Analyses*, in

"Journal of advertising" 50 [1], pp. 81-91.
Jewitt C. (ed.) 2014, *The Routledge Handbook of Multimodal Analysis*, Routledge, London/ New York.
Joshi *et al.* 2017, *Sentiment Resources: Lexicons and Datasets*, in Cambria E. *et al.* (eds.), *A Practical Guide to Sentiment Analysis*, Cham: Springer, pp. 85-106.
Kiritchenko S. *et al.* 2016, *Semeval-2016 task 7: Determining sentiment intensity of English and Arabic phrases*, in *Proceedings of the International Workshop on Semantic Evaluation*, San Diego. Doi:10.18653/v1/S16-1004
Kotras B. 2020, *Opinions that matter: The hybridization of opinion and reputation measurement in social media listening software*, in "Media Culture and Society" 42 [7-8], pp. 1495-1511.
Kress G. 2010, *Multimodality. A social semiotic approach to contemporary communication*, Routledge, London/ New York.
Kress G. and van Leeuwen T. 2020, *Reading images: The grammar of visual design*, Routledge, London/ New York.
Lewandowsky S. *et al.* 2017, *Beyond misinformation: understanding and coping with the 'Post-Truth' era*, in "Journal of Applied Research in Memory and Cognition" 6 [4], pp. 353-369. https://doi.org/10.1016/j.jarmac.2017.07.008.
Liu B. 2017, *Many Facets of Sentiment Analysis*, in Cambria E. *et al.* (eds.), *A Practical Guide to Sentiment Analysis*, Springer, Cham, pp. 11-40.
Liu B. 2012, *Sentiment Analysis and Opinion Mining*, in *Synthesis Lectures on Human Language Technologies*, Morgan & Claypool, Toronto. https://doi.org/10.2200/S00416ED1V01Y201204HLT016.
Liu B. *et al.* 2011, *How publics respond to crisis communication strategies: The interplay of information form and source*, in "Public Relations Review" 37, pp. 345-353.
Majumder N. *et al.* 2018, *Multimodal sentiment analysis using hierarchical fusion with context modelling*, in "Knowledge-Based Systems" 161, pp. 124-133.
Morgan A. and Wilk V. 2021, *Social media users' crisis response: A lexical exploration of social media content in an international sport crisis*, in "Public Relations Review" 47 [4] https://doi.org/10.1016/j.pubrev.2021.102057.
Moschini I. and Sindoni M. G. (eds.) 2022, *Mediation and Multimodal Meaning Making in Digital Environments*, Routledge, London/ New York. https://doi.org/10.4324/9781003225423.
O'Halloran K. L. *et al.* 2014, *Multimodal pragmatics*, in Schneider K. and Barron A. (eds.), *Pragmatics of Discourse*, De Gruyter Mouton, Berlin, pp. 239-268.
Odin R. 2022, *Spaces of Communication: Elements of Semio-Pragmatics*, Amsterdam University Press, Amsterdam.
Palttala P. *et al.* 2012, *Communication gaps in disaster management: Perceptions by experts from governmental and non-governmental organizations*, in "Journal of Contingencies and Crisis Management" 20, pp. 2-12.
Perakakis E. *et al.* 2019, *Social Media Monitoring: An Innovative Intelligent Approach*, in "Designs" 3 [2], pp. 1-12.
Schwaiger J. *et al.* 2021, *UR: SMART–A tool for analyzing social media content*, in "Information Systems and e-Business Management" 19, pp. 1275-1320.
Sindoni M. G. 2013, *Spoken and written discourse in online interactions: A multimodal approach,* Routledge, London/ New York.
Tam L. and J.-N. Kim 2019, *Social media analytics: How they support company public relations*, in "Journal of Business Strategy" 40 [1], pp. 28-34.
Umar A. B. *et al.* 2022, *How Advanced Technological Approaches Are Reshaping Sustainable Social Media Crisis Management and Communication: A Systematic Review*, in "Sustainability" 14 [10], pp. 1-26.
van Noort G. and Willemsen L. 2012, *Online damage control: the effects of proactive versus reactive webcare interventions in consumer-generated and brand generated platforms*, in "Journal of Interactive Marketing" 26 [3], pp. 131-140.
van Zoonen W. and van der Meer T.V 2015, *The importance of source and credibility perception in times of crisis: Crisis communication in a socially mediated era*, in "Journal of Public Relations Research" 27 [5], pp. 371-388.
Ziora L. 2016, *The sentiment analysis as a tool of business analytics in contemporary organizations*, in "Studia Ekonomiczne" 281, pp. 234-241.