FACING #EMERGENCIES
The linguistic role of keywords, hashtags and retweets in communicating critical events

FRANCESCO MELEDANDRI
UNIVERSITY OF BARI

Abstract – In recent years, social media platforms have had a tremendous impact on the online world due to their effectiveness in multimodal communication events (Herring 2001). Social media users benefit from the digital nature of such interactions to gather data for different purposes, including discourse-related ones (Zappavigna 2012). Hashtags, in this sense, have proved to be an effective tool that is used to broaden communication but also to prompt real-life actions, especially to face and manage critical or emergency situations (Olteanu *et al*. 2015). This specific device is more likely to be used effectively on Twitter, a popular micro-blog used as an information aggregator and catalyst for action (Zappavigna 2015). Following previous studies focusing on the same topic (Burnap *et al.* 2014; Hughes, Palen 2009), the paper examines examples of context-based words and/or purpose-specific hashtags to explore their use by different sets of users with diverse intentions and aims. Twitter data are retrieved by means of real-time data mining tools (Brooker *et al.* 2016) to create relevant keyword- or hashtag-based sample corpora dealing with emergency situations (Aug – Sep 2017: two terror attacks and a natural disaster) which have caused remarkable media exposure. Data from such corpora identify some relevant words used in such situations, grouped according to several variables such as event-related, channel-dependent or sentiment-based criteria. Furthermore, an aggregated analysis is carried out in order to retrieve the most common patterns used to highlight performativity, thus emphasising the role of purpose-specific communication. Finally, a comparison of aggregated corpora with different tool-specific features (*retweets*) highlights the importance of such tool-specific devices in magnifying the range of communication effectiveness occurring in a proper 'online discourse community' (Herring 2008).

Keywords: communication; discourse community; social media; hashtags; emergencies**.**

1. Introduction

In recent years, communication in its multi-faceted forms has been experiencing a new upsurge due to the proliferation of favourable conditions. Today, the online digital world stands out as an arena in which anyone has a chance to provide their point of view and to magnify the range of effectiveness of a message. This is due to several reasons, such as the enhancement of digital technologies for a growing audience of potential users (also in terms of different age groups, social and education-related conditions, and geographical areas), but also in terms of the variety of alternatives used in order to share both verbal and non-verbal forms of communication.

The number of Internet users has witnessed a constant growth, with a dramatic rise in recent years. According to statistics, in 1995 only 0.4% of the world population accessed the Net; by the end of 2020 it is likely that 5 billion users could access the Internet, representing almost 63% of the entire world population. At the same time, in the last 8 years online users almost doubled (being 2.5 billion in December 2012);[1] this is

---

[1] Data provided by *Internet World Stats*; https://www.internetworldstats.com/emarketing.htm (11.12.2020).

especially due to the 'break into' new markets such as Asia, which accounts for almost half of the entire world Internet audience.[2] This globalised scenario, in turn, opens a broad discussion questioning the importance of a common language to be used in international and/or universal communication, in which English plays a major role, though there are more users of Asian languages in Eastern areas (and markets) for intralingual communication (Crystal 2011, p. 78). Nevertheless, irrespective of the language used for any interaction, it is important to underline the significance of the medium used for real-time, rapid transmission of ideas as well as its evolution over time, which has allowed the development of new (virtual) areas of contact among people for different aims.

The need for language connections is innate in humans; therefore, the Internet may be considered the ultimate standard for the dissemination of language acts for an innumerable range of purposes (Herring 2015). In this light, this paper focuses on a particular communication need involving time- and space-oriented variables by means of a specific online channel. As a matter of fact, using the Internet – with particular emphasis on specific tools which gather users sharing similar purposes – as a way to route specific information and specialised language, it is possible to see the growth of a common ground with well-defined variables which share the same communicative intention, as this research tries to elucidate.

This contribution introduces empirical research involving the use of a social media platform, Twitter, used to collect data by means of a retrieval of relevant words and tool-specific features. In particular, the context in which the data collection and analysis take place involves the identification of a particular communication aim that sparks a discourse in digital and online environments. In this sense, emergency situations are seen as a fertile ground to assess the level of interaction among different language users which may be driven by various needs, whether the urge to provide emergency-related communication or the necessity to share information by taking advantage of the potential offered by online tools.

First of all, the potential offered by social media in creating communities and discourses is outlined. A description of the essential features of Twitter, an online tool used by a global community to share short public texts, is then offered. The case study involves the creation of some sub-corpora resulting from the retrieval of relevant words after emergency-based situations which took place in 2017, triggering significant media exposure in both traditional and online environments. These events refer to the terrorist attack occurred in Barcelona, Spain, on August 17, 2017 (for which the term *Ramblas* and *Rambles* represent two distinct data collection); the arrival of a Category-5 Hurricane named "Irma" in Florida, USA, in early September 2017; and the train bombing at Parsons Green underground station, London, on September 15, 2017. The compilation and analysis of corpora associated with stand-alone events allow the identification of significant words in each corpus by means of their (lemmatised) frequency of use, which could be gathered according to different functions of language. In this sense, the identification of some collocates associated with some relevant tokens may be helpful in defining the event (i.e. a terrorist attack or a natural disaster) more in detail. At the same time, an aggregate analysis, by combining all data deriving from the Twitter corpora related to the four events, may lead to the identification of recurrent patterns of use in emergency or crisis-related situations irrespective of the specific event involved in its divulgation. For this purpose, some colligational patterns can be identified in order to highlight the importance

---

[2] Internet Users in the World by regions – June 30, 2018, *Internet World Stats,* available online at: https://www.internetworldstats.com/stats.htm (11.12.2020).

Facing #emergencies. The linguistic role of keywords, hashtags and retweets in communicating critical events

293

of some verbal processes in communicating emergencies. All sub-corpora are analysed as both standalone and aggregated cases in order to compare the use of interactions in emergency situations according to some tool-specific features such as *hashtags* and *retweets*. Such tools were chosen since they lend themselves well to the communicative purpose chosen as the object of this study, that is the analysis of emergencies and emergency-related communication, which require a rapid and immediate response even in verbal and non-verbal terms.

2. Online communities and Social Media: discourse-based interactions for specific purposes

As previously stated, the potential offered by new and refined means of communication allows the transmission of messages of a various nature and for a diverse range of purposes. This form of interaction can be encompassed in a phrase and a concept popularised by Thurlow *et al.* (2004) when referring to *Computer-Mediated Communication*, or CMC. According to these scholars, communication has also been revolutionised in terms of cultural rebirth thanks to the advent and popularisation of digital devices, in particular the Internet, which make it possible to interact in real time, hence overcoming any constraints of time and space. This, in turn, implies a deep change in social interaction, explained as "how *identities, relationships* and *communities* are being changed or influenced by the Internet" (my emphasis) (Thurlow *et al.* 2004, p. 2). Moreover, the Internet has evolved in its infrastructure technique, resulting in an enhanced transmission of messages by virtue of which multimodality – or "any text whose meanings are realized through more than one semiotic code" (Kress, van Leeuwen 2006, p. 177) – has become a well-established reality. This has led to the diffusion of messages that mix different sign systems via different mediums and on a global scale, thus establishing system-user and user-user communication/genres (Jovanovic, van Leeuwen 2018), or new connections in terms of digital technology from the technical point of view, and social life and culture on the 'human-oriented' side (Kern, Develotte 2018).

From a linguistic point of view, the notion of communication applied to the use and the exploitation of digitally computerised tools can be further redefined in terms of *Computer-mediated Discourse* (CMD). This concept, theorised by Herring, draws on CMC but with a specific focus on the purpose of the message and the interaction among participants, or the medium meant as a genre rather than its technical infrastructure (once again, the latter being a synonym for *multimodal* rather than *multimedia*). As a matter of fact, CMD may be re-interpreted in the light of Baron's words, when she underlined that "Computer-Mediated communication [is] a force in language change" only if it is accepted that computers, though using their own syntax and language (in another form), become "*conduits* of natural language" (my emphasis) (Baron 1984, p. 119). CMD is shaped around some variables that aim to spread messages according to different forms of interaction, involving the definition and the role of communities (with the introduction of the notion of *virtual communities*; Herring 2008) by means of their interaction(s) in linguistic terms and in social implications (Herring 2001).

Considering the above-mentioned framework, the identification of discourse communities may be implemented also for CMD purposes or, put differently, that language use on the Internet may lead to the identification of specific discourse communities. This statement may conflict with some views according to which the Internet, by its own nature as a 'free', unregulated medium ("a powerful arena for

language") with unpredictable effects on language change may not be considered the ideal space for building and identifying discourse communities (Shohamy 2006, p. 128). At the same time, the emergence of the Internet as a language and communication tool on a global scale confirmed the potential of this device in the recognition of communities in which different forms of discourse-oriented samples are found. Moreover, the above-mentioned social implications can be easily retrieved in CMD and, more in general, in online communities, provided that some specific criteria are identified in the limitless ocean of scattered information found on the Internet. When recalling some deep-rooted definitions of discourse communities, similar traits can be retrieved in this new forms of communication for specific purposes and participants. Herzberg emphasizes the role of discourse that "operates within conventions defined by communities, be they academic disciplines or *social groups*" (my emphasis) (Herzberg 1986, p. 1 in Swales 1990, p. 21), thus legitimising both the conventional and framed forms of languages and the social-oriented ones. Swales (2011), in turn, focuses on the reasons for joining and/or being absorbed by discourse communities thanks to the use of common goal-oriented forms of language: "A discourse community recruits its members by *persuasion*, training or relevant qualification […] an archetypal discourse community tends to be a *Specific Interest Group*" (p. 471). In addition, specific aggregations of digital, online or Web texts and/or communication instances are characterised by unique features that develop stand-alone 'genres of the Internet', thus strengthening the paradigm of online discourse communities meant for specific purposes, with a definite use of language (Giltrow, Stein 2009, pp. 1-2) and acting "within a particular socially constructed discourse" (Barton, Lee 2013, p. 32).

Bearing these variables in mind, it is important to underline that the Internet, meant as a general infrastructure, cannot be considered a proper channel to look for examples of discourse community. Only by providing some distinctive CMD variables involving the channels of interaction, the language used and its social implications, some genres prove to be fruitful examples of online discourse communities. In this respect, a new trend is emerging: within the boundless corpus of interactions, *Social Media platforms* emerged as new spaces of linguistic debate and subsequent potential change. Kim and Vorobel define Social media as "online discourse practices [that] have challenged traditional ways of understanding discourse communities" (Kim, Vorobel 2017, p. 274); therefore, they represent a breakthrough in overcoming traditional (or academic) frameworks since they show high degrees of intertextuality and multimodality. Such features fit into a broader "Internet discourse" which is characterised by hybridity aimed at providing "wider linking across domains and types" (Kim, Vorobel 2017, p. 274). After all, when applying Swales' (1990) features for the identification of discourse communities to the new paradigm offered by social media-based discourse, a perfect match can be found. Goals, intercommunication, participation, genres, lexis and expertise can be retrieved in online language acts even by implying new variables, such as the proficiency needed in order to use specific digital and online tools.

Social media are the result of a technical evolution in the infrastructure of the Internet, the latter starting from a unidirectional flow of contents and language (a one-to-many or a few-to-many interaction, with early websites proposing contents to Internet users) to a multi-directional communication, with non-institutional users playing a decisive role in spreading language contents and creating related communities. This approach, called Read/Write Web or Web 2.0, is characterised by the proliferation of the so-called *user-generated contents* (Kaplan, Haenlein 2010)*,* the latter being marked by a reciprocity between reading and writing, and play "an integral role in comprehension

Facing #emergencies. The linguistic role of keywords, hashtags and retweets in communicating critical events

295

when writing and communicating on the Internet" (Johnson 2014, p. 10). Consequently, discourse-related perspectives should be considered by taking the relevance of Discourse 2.0 into account; thus, CMD shapes around any new form of CMC-based instances of language (Herring, Androutsopoulos 2015).

In Social media platforms, the idea of *community* (and its related forms of discourse) is very strong to the point where it overcomes and overshadows the idea of *identity*, or the association of a precise addresser to any message conveyed via these platforms. As a matter of fact, some social networking sites do not require a compulsory indication of personal details; therefore, they can be divided into identity-based vs. non identity-based social media. This implies that the content or the message provided, framed in the appropriate context(s), are more relevant than the addresser of the communication act. In this sense and recalling Firth's theories which combined personality and language rather than persons, Martin agrees that "we're not looking at individuals interacting in groups but rather at persons and personalities communing in discourse" (Martin 2009, p. 563), thus confirming that interactions in social media are important especially in conveying instances of language according to different variables. The non-compulsory presence of identifiable addressers, of course, is not synonym for low-quality content: indeed, non identity-based social networking sites or communities are able to provide high quality contents based on the relevance of the message and its communicative purpose for and within the community itself (Agichtein *et al*. 2008). At the same time, such messages provide personal experience both in technical terms (an experience of use on social media, with different levels of scalability) and, most of all, in terms of the kind of experience (personal activities leading to emotional-based judgements or evaluative statements) one aims at sharing (Jikoun *et al.* 2010).

3. Social media platforms for discourse-oriented language practices: Twitter

Social media differ according to the communication-based aims they seek to attain. Though they can be comparable in technical terms – "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content" (Kaplan, Haenlein 2010, p. 61) – they have their own rules and purposes, according to the kind of audience or community they would gather and develop. Social media platforms can be divided into two broad categories: the first one aims to create specific communities, gathered according to some variables such as geographical areas, particular interests and specialised fields of discussion. In this category it is possible to include VKontakte,[3] Weibo[4] (two general social networking sites meant for defined areas or nations such as Russia and China), Instagram[5] (a photo and video sharing site), Linkedin[6] (Business and professional networking), Academia.edu[7] (a site for academics and researchers), and aNobii[8] (meant for book lovers). The second category includes social networking sites meant for general

---

[3] https://vk.com/ (11.12.2020).
[4] https://www.weibo.com/login.php (11.12.2020).
[5] https://www.instagram.com/ (11.12.2020).
[6] https://www.linkedin.com/ (11.12.2020).
[7] https://www.academia.edu/ (11.12.2020).
[8] https://www.anobii.com/ (11.12.2020).

purposes, the aim of which is to gather a broad community with no potential limitations (with the exception of the age limit, pursuant to national privacy laws or international regulations such as the recent General Data Protection Regulation, or GDPR)[9] and to share wide-ranging contents. The most popular ones in this sense are Facebook[10] and Twitter.[11] Since social media are all becoming attractive for users, they represent a very interesting field of research in order to analyse language according to different variables (Bredl *et al.* 2014). Twitter, in particular, has shown itself to be useful and somehow preferable in terms of Discourse Community Analysis due to some language practices and the use of tool-specific features (Zappavigna 2012).

Twitter is a popular micro-blogging platform founded in 2006 by Jack Dorsey and characterised by a growing global audience. The company has a revenue of $3.45 billion, and the number of monthly active users is 330 million (as of December 31, 2019). The number of tweets sent per day is more than 550 million,[12] with these data providing indications of the potential offered by this platform for linguistic analysis.

In terms of language, Twitter differentiates itself from other competitors for a series of distinctive features. The most emblematic is the fact that messages, called *Tweets*, could not exceed a preset number of characters (O'Reilly, Milstein 2012, p. 18). Since its early development, this number was set at 140 characters; recently, the length of a tweet has been extended to 280 in order to 'stretch' users' ideas and thoughts.[13] Interestingly, notwithstanding this amplification of 'co-textual virtual space' for users, the number of tweets exceeding the previous limit of 140 characters was only 5% higher, with 2% of them exceeding the threshold of 190 characters.[14] This means that users have been using this platform according to a pre-established, distinctive feature of this device resulting in a direct, concise and effective language. In other words, Twitter is different from other social networking competitors thanks to this co-textual, density-based rule that users *want* to use this way.[15] The conciseness of Twitter has some repercussions on the (discourse) community using this device. First of all, Twitter may favour quality contents since users use the co-textual limit in order to provide direct and effective communication to their audience; what is more, the use of tool-specific devices on its platform somehow contains its community, thus refining it in terms of expertise. Following this assumption, the fact that Twitter is less popular than other general social networking sites such as Facebook may favour the specific nature of the contents provided. Users have a decisive role in leading the platform towards this quality-based language use, and this is due to the kind of content provided via Twitter: evidence of this can be seen in the change of Twitter's prompt (the message provided to their users when they start tweeting) from *What are you doing*? to *What's happening?*. This change in the prompt shifts users' attitude and

---

[9] Age limit restrictions are specified in Article 8 of the GDPR. https://gdpr-info.eu/art-8-gdpr/ (11.12.2020).

[10] https://www.facebook.com/ (11.12.2020).

[11] https://twitter.com/ (11.12.2020).

[12] *Twitter Annual Report 2019*. https://s22.q4cdn.com/826641620/files/doc_financials/ar/2018/AnnualReport2018.pdf (11.12.2020).

[13] Rosen A. 2017, *Giving you more characters to express yourself*, Official Twitter Blog, September 26, 2017. https://blog.twitter.com/official/en_us/topics/product/2017/Giving-you-more-characters-to-express-yourself.html (11.12.2020).

[14] Rosen A. 2017, *Tweeting Made Easier*, Official Twitter Blog, November 7, 2017. https://blog.twitter.com/official/en_us/topics/product/2017/tweetingmadeeasier.html (11.12.2020).

[15] Lu L. 2017, *Experimenting To Solve Cramming*, Engineering Twitter Blog, November 7, 2017. https://blog.twitter.com/engineering/en_us/topics/insights/2017/Experimenting-To-Solve-Cramming.html (11.12.2020).

Facing #emergencies. The linguistic role of keywords, hashtags and retweets in communicating critical events

297

focus to the contents to be provided from a personal perspective to general, quality-based information. Indeed, Twitter defines itself as a device where people describe "what's happening in the world and what people are talking about right now".[16]

Discourse, then, is constructed on the basis of some peculiar features. On Twitter, communication is based on the Following-Follower criterion: users decide which users are worth following and vice versa, so they will get a number of followers who receive notifications of the tweets sent. This creates sub-communities since users are more likely to follow their interests and to gather according to common fields.

However, the most striking features provided by Twitter, which will represent a key instrument for the case study analysed in this paper, are two device-specific tools: Hashtags and Retweets.

Hashtags represent a technical use of language that marks a difference in making Twitter a discourse-oriented tool. Technically speaking, they are terms or expressions prefixed by the hash (#) symbol which could be found in any digital device (from personal computers to smartphones and even on 'traditional' phones) and whose function is to categorise messages (O'Reilly, Milstein 2012, p. 18). In this way, any topic or keyword may be easily gathered in a single list of tweets containing a specific hashtag, thus providing a classification in semantic terms, which is helpful in reconstructing discourse within a sub-community of Twitter users (Dickinson 2013). Hashtags can be conceived within a twofold perspective: the first involves a technical-oriented vision, since it relates to the use of a specific digital tool used in ICT terms. In this sense, hashtags are the result of a non-visible syntax used by programming language that enacts the creation of this feature and its related use by Twitter members. This use of additional data, called *metadata*, is used to organise information. On the other hand, seen from a language perspective, these metadata "support social relations". They fulfil this dual perspective because they have a textual function with deep experiential and interpersonal implications (Zappavigna 2014). Hashtags, then, prove to be an effective tool for media discourse, which can be accomplished only by setting out some specific criteria. A fundamental condition is represented by searchability, or the possibility to retrieve contents on the basis of well-defined queries – thus marking the idea of "aboutness" of communication (Kehoe, Gee 2011). For this reason, the option offered by hashtags is part of what Zappavigna calls the "searchable talk", a form of aggregation of contents which "affords the possibility of new forms of social bonding" (Zappavigna 2015, p. 290) but also exists as "part of the linguistic structure and discourse semantics and also as metadata" (Zappavigna 2015, p. 275).

Another effective tool on Twitter is represented by *Retweets*, or RTs, that is a one-click feature that allows users to republish contents by other users. By retweeting, users usually do not alter the original message provided by the original addresser, but they post it as one of 'their' contents, so that the information is available in the tweet stream of the followers who have clicked on RT. According to Boyd *et al*. (2010, p. 1), "this practice contributes to a conversational ecology in which conversations are composed of a public interplay of voices that give rise to an emotional sense of shared conversational context". In this way, a message is amplified and the reach of a message is potentially unlimited, since it can be made available to an indefinite number of users beyond the criterion *following-followers*. The 'power' of Retweets depends on several variables such as the popularity of the sender's account but also the kind of information provided, two criteria

---

[16] This is the initial message in Twitter's *About Us* section. https://about.twitter.com/en_us.html (11.12.2020).

that confirm once again that quality-based contents can also be fostered by means of these device-specific features.

4. When Social Media discourse meets social causes: the case of emergencies and crises

Social Media represent a virtual space where effective language and discourse can be retrieved. Given its popular nature and the variety of fields involved in the 'searchable talk', each tweet is potentially meant for a specific and somehow imagined audience that a user is addressing (Marwick, boyd 2010), for instance by using unique and semantically relevant hashtags (Bruns, Burgess 2011). An interesting case could be represented by the use of Social Media as a channel to convey rapid messages for an audience that should be as broad as possible; in this sense, social causes could provide interesting insights for language use via digital and online tools. Among the diversified range of social media users (with non-identification and diversity being two 'pillars' of public, non-identity-based social networking sites) (Leppänen *et al.* 2017) an essential distinction would have to be made between public users, that is accounts created in the name of governmental or commercial institutions, and private users, that is anyone with no affiliation with such companies or institutions. Private users are more numerous than public ones, and the importance of such users is growing due to the increasing significance of "new spaces of power for citizenry engagement" provided by Social Media (KhosraviNik 2017, p. 583). However, public users play an important role in aggregating followers in order to create communities, especially according to geographical parameters. For these reasons, the communication of institutional messages in crisis conditions has emerged as a growing circumstance useful for observing the development of language use and communication patterns (Bruns, Stieglitz 2012). By intertwining the messages provided by institutions and private users, the management of crisis can lead to positive results in highlighting topic-related contents which prove to be favourable and 'pushed' by the online community (unlike topic-related tweets containing rumours and/or inaccurate information) (Mendoza *et al.* 2010). In this way, a many-to-many interaction and controlling of the communication provided by different sources and for different purposes acts as a controlled discourse sample within the realm of the so-called *Big Data* (Bruns, Liang 2012).

5. A corpus-based analysis of emergency communication via Social Media: four case studies

In the light of the possibilities offered by Social Media, it is possible to collect and gather a set of data from social networking sites in order to analyse the behaviour of users in terms of discourse in the communication of emergency and crisis situations. Due to its features, Twitter can be exploited for different purposes and by diversified communities, which gather on the occasion of such generalised events that involve whole societies (Hughes, Palen 2009). Starting from this perspective, data can be collected to create language corpora to be analysed in terms of specialised patterns which may be instantiated both in institutional profiles (driving the tool-specific language for a uniform language-oriented management of the crisis event) and among private users who may provide (quality) useful information acknowledged by the online community.

Facing #emergencies. The linguistic role of keywords, hashtags and retweets in communicating critical events

299

5.1. Method

5.1.1. Definition of case studies

The case studies analysed in this paper refer to three major events which happened in the timespan between August 2017–September 2017. Though referring to specific areas, these events had international repercussions due to their significance, thus generating reactions in both 'traditional' communities and virtual ones. The latter is particularly important for this study, since the language used to react, in terms of expressing users' thoughts and helping institutions and the community, represents the main focus of this analysis.

Following the criteria set out by Olteanu *et al.* (2015) in a similar work, it is important to define what kind of crisis or emergency situations could be analysed. Hazard type, temporal development and geographic spread are three variables to define emergencies. Hazard type includes natural and human-induced events; some sub-categories can also be defined (meteorological, hydrological, geophysical, etc. for natural events; intentional or accidental for human-induced ones). On the other hand, temporal development identifies a timespan in which the event has its peak before or after the observation period. This lapse of time depends on whether the event has instantaneous origin (e.g., an earthquake or a shooting), so that the temporal development to be observed happens after the event, or is a progressive one (e.g., an hurricane or a planned evacuation) with the temporal development involving both the stages before and after the event itself (Olteanu *et al.* 2015, p. 995). Geographic spread (focalised vs. diffused area of the event) was not taken into consideration in the analysis of tweets, though all data have been collected in English even in case of events which occurred outside English-speaking countries.

The crisis events involved in the analysis are shown in Table 1.

NAME OF EVENT	HAZARD TYPE	TEMPORAL DEVELOPMENT	DATA COLLECTION TIMESPAN	LOCATION
Ramblas/Rambles Attack	Human-induced	Instantaneous	August 17-23, 2017	Barcelona, Spain
Hurricane Irma	Natural	Progressive	September 2-11, 2017	Florida, USA
Parsons Green terror threat	Human-induced	Instantaneous	September 15-16, 2017	London, UK

Table 1
List of the crisis/emergency events analysed.

5.1.2. Data collection

In order to collect data from Twitter, it is necessary to draw on the platform public data access. In this sense, Twitter allows users access to its data via its *Application Programming Interface (API)*,[17] so that data could be retrieved by specifying definite

---

[17] Further info at: *Developer Twitter Docs* – https://developer.twitter.com/en/docs (11.12.2020).

criteria following a default ICT syntax (Bruns, Liang 2012). An API consumer key for developers is needed in order to engage with the Twitter platform. It can be obtained after creating a Twitter account[18] and developing a Twitter App.[19] This procedure allows users-developers to interface with the platform and to search for Twitter data (Tweets) on the basis of different queries.

5.1.3. Retrieval tool: Chorus Tweetcatcher Desktop

In order to gather data from Twitter, a specific software was used. Chorus Tweetcatcher Desktop (CTD) is a free Social Media analytics tool and part of the Chorus project (Brooker *et al.* 2016). This tool is used in order to retrieve Twitter data according to simple or complex queries. In order to work, the software needs the above-mentioned credentials represented by the Twitter API consumer key and consumer secret (a password).[20] Twitter data "lend itself to the exploration of 'topics'" (Brooker *et al.* 2016, p. 5). By entering queries and selecting the timespan of the retrieval, the software is able to search for tweets containing hashtags and/or semantically dense keywords related to any event, in this case represented by emergency-oriented words. In this way, an almost real-time retrieval of contents in the *Twittersphere* could be obtained, thus allowing a pragmatic analysis of the data collected (Hoffmann, Bublitz 2017). The retrieval process leads to the creation of Social Media data that can be analysed according to a corpus-based perspective. Some limitations to this methodology can be represented by the fact that restrictions are imposed by Twitter in the retrieval of data; in this sense, only a 7-day backwards retrieval is allowed, and this may lead to the recovery of incomplete data (i.e., in progressive emergency events). What is more, some tool-specific language instances (represented by the retrieval of technical information such as URLs or part of them) may create noisy data especially in the stage of corpus analysis.

5.1.4. Creation of emergency-based corpora of Tweets

Data obtained from Tweetcatcher were filtered by excluding non-essential metadata. For this purpose, only the actual message sent by the user, that is the *Tweet*, was included in the corpus. Tweetcatcher saves data in a tab-delimited format. This allows for the isolation of single categories (in this case, the category "Tweet") which can be read through other applications such as Microsoft Excel or text-editors. The isolation of language-based occurrences allows for the creation of a corpus, since data have been saved as plain text file and imported into corpus analysis tools. Corpus analysis can be understood as the investigation of "whole datasets as an 'information space' in which semantic features (words, hashtags, etc.) intersect in potentially interesting ways, […with] [r]esearchers […carrying out] the exploration of topical structures emerging from the entire body of data. […] (Brooker *et al.* 2016, p. 5). The analysis was carried out by a manual inspection of the most recurrent words or hashtags used, following or in association with the event (Burnap *et al.* 2014), but the corpus-based analysis allows for a deeper investigation of the whole language set in terms of discourse analysis.

---

[18] https://twitter.com/i/flow/signup (11.12.2020).
[19] https://developer.twitter.com/apps (11.12.2020).
[20] The software and user guides are available at: http://chorusanalytics.co.uk/ (11.12.2020).

5.1.5. Corpus compilation

Any corpora created starting from a retrieval of Tweets should be considered within a twofold perspective. On the one hand, a corpus collects a wide range of single communication acts, represented by Tweets, which constitutes a sample of language use for specific, though diversified, contexts and purposes. At the same time, given the particular nature of the medium used for this communication, the possibility of reiterated language acts is rather high due to some tool-specific possibilities. Given these assumptions, corpora were created independently for single analyses and then aggregated in order to provide collective data (a sample corpus of general emergencies). For each corpus deriving from the retrieval via CTD, a sub-corpus excluding Retweets (RTs) has been created in order to obtain standalone, non-repeated data. As for the Ramblas/Rambles event, two corpora have been created on the basis of two different retrieval criteria: one involved the hashtag/keyword "Ramblas" for the purpose of identifying the precise place of the terror attack carried out in Barcelona on August 17, 2017. Another was based on the query "Rambles" which identifies the same name in Catalan. Keywords can be divided into sub-categories that can be associated with the emergency they refer to; at the same time, the sub-categories represent a way to distinguish different approaches to the same event. The sub-categories identified are event-related tokens (words that describe the event), geographical-related tokens (words used to designate the event in spatial terms), information-related tokens (words used to refer to media-oriented lexis, provided that media outlets are usually considered a source of reliable information), sentiment-related tokens (words used by users in order to express their feelings and/or positions towards the event) and channel-dependent tokens (in this case, popular hashtags). 'Pure' corpora and corpora with RTs have been compared in terms of frequency of use. Table 2 and Table 3 show the size of the corpora used for this study, specifying comprehensive data and pure, one-time tweet corpora, respectively.

CORPUS NAME/KEYWORD	NO. OF TWEETS	CORPUS SIZE (TOKENS)	WORD TYPES
RAMBLAS	150k+ tweets	3M+	53k+
RAMBLES	14.8k+ tweets	295k+	9.7k+
#HURRICANEIRMA	186k+ tweets	3.9M+	54k+
#PARSONSGREEN	160k+ tweets	3.3M+	37.6k+

Table 2
Size of corpora in terms of number of tweets, number of tokens, number of word types.

CORPUS NAME/KEYWORD	NO. OF TWEETS	CORPUS SIZE (TOKENS)	WORD TYPES
RAMBLAS	30k+ tweets	580k+	48k+
RAMBLES	1.7k+ tweets	30.4k+	6.5k+
#HURRICANEIRMA	12.1k+ tweets	427k+	46k+
#PARSONSGREEN	22k+ tweets	400k+	32.8k+

Table 3
Size of corpora without Retweets (RT) in terms of number of tweets, number of tokens, number of word types.

The size of the general, aggregated corpus is shown in Table 4. Data refer to the corpus without retweets, thus showing one-time tweets only.

CORPUS NAME	NO. OF TWEETS	CORPUS SIZE (TOKENS)	WORD TYPES
Aggregated corpus	65.8k+ tweets	1.43M+	133.3k+

Table 4
Size of the aggregated corpus in terms of number of tweets, number of tokens, number of word types.

The size of the aggregated corpus including Retweets is shown in Table 5.

CORPUS NAME	NO. OF TWEETS	CORPUS SIZE (TOKENS)	WORD TYPES
Aggregated corpus + RT	510.8k+ tweets	10.6 M+	154.3k+/129k+*

Table 5
Size of the aggregated corpus with Retweets in terms of number of tweets, number of tokens, number of word types.
* The first figure indicates the overall number of word types. The second figure indicates the number of word types after implementing a stoplist with 'noisy' tokens (such as URLs).

6. Results

6.1. Use of relevant keywords for emergency-related communication

The analysis of the corpora shows some interesting data in terms of the use of keywords deriving from the communication of emergencies by both public institutions and private users. What follows is a list of field-specific terms found in the corpora, which show a high frequency of use, retrieved by means of lexical rather than statistical relevance. The reason for this choice lies in the nature of the retrieval; the data were collected by focusing on the presence of a given lexical item in a tweet which referred to the nature or name of the event, or the geographical area in which the emergency occurs (*Corpus Name* in Table 3). At the same time, the list of relevant words deriving from the analysis of the different sub-corpora is based on the frequency of use of tokens or lemmatised items found in each sub-corpus. Data refer to corpora without Retweets, therefore representing independent and non-repeated messages related to the events. Table 6a, 6b, 6c and 6d summarise the keywords found in each corpus, indicated as lemmatised items and arranged in alphabetical (not frequency) order, characterised by a minimum frequency of 200, 10, 150 and 150, respectively, that is a proportional threshold to the general size of the corpora.

CORPUS NAME	RAMBLAS (FREQUENCY: >200)
Event-related tokens	area, arrest*, attack*, carnage, crash*, crowd*, dead*, hits, horror, incident, injur*, kill*, missing, police, suspect*, terroris*, tourist*, truck, van, victim*
Geographical-related tokens	Barcelona, Cambrils, Catalan, city, district, las, ramblas, Spain, Spanish, street
Information-related tokens	BBC, breaking, CNN, confirm, fatalities, media, news, report*, update*

Facing #emergencies. The linguistic role of keywords, hashtags and retweets in communicating critical events

303

Sentiment-related tokens	defiance, famil*, fight, ISIS, Islamic, jihadi, love, stop, thoughts
Channel-dependent tokens (Hashtags)	#barcelonaattack, #lasramblas

Table 6a
Main Keywords and hashtags in the Ramblas Corpus.

CORPUS NAME	RAMBLES (FREQUENCY: >10)
Event-related tokens	attack, emergency, incident*, injur*, mossos, police, rambles, terror*, van, victims
Geographical-related tokens	Barcelona, Cambrils, Cat, Catalan, Catalonia, Catalunya, city, rambla, ramblas, Spain, Spanish
Information-related tokens	live, latest, news, update, video
Sentiment-related tokens	family, God, love, please, prayers, solidarity, stop, thank, thoughts, walk*
Channel-dependent tokens (Hashtags)	#barcelonaattack, #emergenciescat, #lasramblas, #totsombarcelona

Table 6b
Main Keywords and hashtags in the Rambles Corpus.

CORPUS NAME	HURRICANE IRMA (FREQUENCY: >150)
Event-related tokens	category, destruction, devastation, disaster, emergency, evacuat*, hurricane, impact, Irma, power, residents, safe*, shelter*, storm, water, wind*, weather
Geographical-related tokens	Atlantic, Barbuda, beach, Caribbean, coast, Cuba, Florida, Floridians, Georgia, Haiti, island*, key*, Miami, Ocean *, south, Tampa
Information-related tokens	Breaking, CNN, information, latest, live, news, update*
Sentiment-related tokens	Family, help, please, pray*, thoughts, victims,
Channel-dependent tokens (Hashtags)	#floridakeys, #irmahurricane, #hurricaneirma, #hurricaneirmatracking, #keywest, #periscope, #stormwatch

Table 6c
Main Keywords and hashtags in the Hurricane Irma Corpus.

CORPUS NAME	PARSONS GREEN (FREQUENCY: >150)
Event-related tokens	Arrest*, attack*, blast, bomb, critical, emergency, explosion, incident, injur*, level, police, security, terror*, threat, train, victims
Geographical-related tokens	London, station, tube, underground, UK
Information-related tokens	breaking, information, news
Sentiment-related tokens	ISIS, Islam, love, pray, sad, stop, thoughts, victims
Channel-dependent tokens (Hashtags)	#londonunderground #metpoliceuk #Parsonsgreen #sadiqkhan

Table 6d
Main Keywords and hashtags in the Parsons Green Corpus.

6.2. Aggregated analysis: recurrent patterns for emergency performativity

Though they refer to specific emergency events, each corpus has its own peculiarities in terms of the use of the lexis, the latter depending on the nature of the crisis-related occurrence. At the same time, the narration of these events by means of tweets shows some similar features in all corpora, showing examples of descriptive, informative or emotive language acts (Reiß, Vermeer 2014, p. 182).

Some verbal processes and the use of other Part-Of-Speech (PoS) elements can be analysed in the light of emergency-related instances of language. These elements prove to be very important in communicating some essential information associated with critical events for different purposes. An example is represented by the verb *to avoid*, which leads to the patterns *V+obj.* (*avoid the area, avoid las ramblas, avoid #parsonsgreen, avoid the collapse of phone*) or *sth + to avoid* (*people are advised to avoid, using social networks to avoid*) used in order to provide the population with some useful advice for managing emergencies, thus being instructive messages. Similarly, the verb *to stay* shows a high number of occurrences of the phrases *stay safe, stay away* or *stay inside* used in order to ask people to be cautious. Emotion-induced tweets highlight the presence of the phrase *stay strong*, a way for users to show their support for the emergency events (both natural and human-induced). Informative communication is shown by the high incidence of the verb *to confirm*, a verb used by institutions or media in order to provide official information about the number of casualties or updates in the investigation process. Here, patterns such as *confirmed dead* (with the left co-text showing different figures according to the specific event or the update stage), *confirm fatalities* or *police confirm, police have confirmed, family confirms* all prove the use of this channel of communication in order to provide official information related to emergency or crisis situations.

It is interesting to note the incidence of the verb *to share* in the aggregate corpus to verify some common patterns and their related function. If examples such as *share their #hurricaneirma stories* emphasize the role played by social media for entertainment purposes, other examples such as *sharing cat pictures* or *don't share* seem more out-of-context patterns when referred to such events. Though *sharing cat* may refer to the code used for Catalonia (with examples in the corpus retrieved from Ramblas or Rambles

Facing #emergencies. The linguistic role of keywords, hashtags and retweets in communicating critical events

305

corpora), the analysis of contexts including these patterns refers to an initiative spread by Twitter users of posting pictures of cats in their tweets in order not to divulge information that could be used for the investigation, but also in a form of respect to the people involved in a terror attack.[21] Cats, in this case, are used as part of the so-called *meme culture*, or the viral use of certain images and representations for humorous purposes. On the other hand, the recurrence of *don't share* follows the indications by authorities not to post distressing content (especially in human-induced events).[22] A similar case is represented by *don't spread*; at the same time, this verb has also been used for sentiment-based opinions as in the case of the collocation *spreading hope*, which is part of an encouraging, emotional communication.

Finally, considering the nature of real-time communication which can be enacted via social media platforms, a particular case is represented by the term *information*, whose collocates include event-related words such as *[about the] attack, incidents*, or more specifically *#parsonsgreen* and *#hurricaneirma*. This use of the medium for informational purposes is confirmed by *if you need any information*, which is presumably originated by institutional accounts towards the online community, in a one-to-many communication process, and the semantically complementary collocate *if you have information*. Here, institutions call for the help of citizens in helping them for investigations in case of useful information or media. This communication may clash with the above-mentioned request of avoiding the sharing of multimedia contents; however, by looking closely at the actual co-texts of this communication, it appears that anyone able to provide useful information is asked to contact authorities by means of 'traditional', private channels, thus confirming that the potential of real-time information used to manage emergency situations sometimes requires users to abide by ethical principles by avoiding mass, public and many-to-many interaction channels.

6.3. The power of RTs: amplifying the reach of useful information for a global audience

The purpose of the "Retweet" function is to amplify the range of a given message, especially in case of useful or popular information on a potentially limitless scale. For this reason, an aggregated analysis of the corpus including Retweets is useful for the purposes of this study in order to assess the degree of propagation of patterns and the resulting language acts identified for this particular type of events.

Some interesting, quantitative data emerge from the comparison of the incidence of the collocations outlined in the previous section with the corpus containing RTs. As for the colligation *avoid + obj*, the phrase *avoid the collapse of phone* has a frequency of 3940 occurrences, while in the one-tweet corpus it was only 15 (A). Similarly, *avoid the area* accounts for 1214 occurrences compared with 62 in the other corpus (B). More general requests such as *please avoid* confirm the collocation is used to spread utility-based communication (594 in the RT corpus; 21 in the aggregated, non-RT one) (C). As for the verb *to stay*, the most frequent collocation is *stay safe*, with RTs playing a role in magnifying its occurrence (3582 vs. 977) (D). In view of the information-related dimension of Social Media, patterns such as *confirm fatalities* (E) or *confirmed dead* (F)

[21] Zorthian J. 2017, *Here's Why People Are Posting Cats in Response to the Barcelona Attack*, in "Time Magazine". http://time.com/4905356/barcelona-terror-attack-cat-posts/ (11.12.2020).
[22] The tweet by the Spanish police during in the immediate aftermath of the Barcelona attack is available online at: https://twitter.com/policia/status/898209070993338368 (11.12.2020).

show a considerable increase in terms of occurrences (1813 vs. 46 and 1273 vs. 459, respectively). Considering the left co-text of the verb *to confirm*, the pattern *police confirm* peaks at 2939 occurrences against 229 examples in the one-time corpus (G). The verb *to share* is very likely to be found in social media communication, since it is used when users want to spread messages to their network(s). In the corpus, it is interesting to note that a pattern such as *don't share* reveals a considerable propagation of occurrences (953 vs. 180) (H), while others remain somehow isolated linguistic phenomena with a scarce consideration by the online community, with *sharing cat pictures* showing no significant difference in the number of occurrences in both corpora. Similarly, *don't spread* proves to be an effective message, showing a marked increase of occurrences (601 vs. 14) (I) with its variation *do not spread* showing similar results (144 vs. 3) (J). The examples involving the node *information* confirm the relevance of RTs: patterns such as *if you need any information* (K) or *if you have information* (L) are messages that the online community considers a relevant language act; by retweeting it, an important message that could be used to manage or solve emergencies could get a boundless echo (3388 vs. 25 and 1062 vs. 11, respectively).

Giving the discrepancy in terms of corpus size when comparing the frequency of these occurrences, raw frequencies do not provide an objective ratio in defining the relevance of specific patterns. In the assessment of quantitative results from corpus analyses, normalisation of data (McEnery, Hardie 2012, p. 51) needs to be implemented in order to confirm the effectiveness of Retweets in highlighting the relevance of phrases typically associated with emergency situations. Table 7 shows the normalised frequency (per one-million tokens) of the above-mentioned patterns in both aggregated corpora without and with Retweets, respectively. These figures show a higher frequency in the aggregated corpus that includes Retweets, with some of them indicating a marked difference (e.g., A, B, and L).

PATTERN	RAW FREQUENCY - AGGREGATED CORPUS	NORMALISED FREQUENCY (PER 1M TOKENS) - AGGREGATED CORPUS	RAW FREQUENCY - AGGREGATED CORPUS + RT	NORMALISED FREQUENCY (PER 1M TOKENS) - AGGREGATED CORPUS + RT
A	15	**1.04**	3940	**371.69**
B	62	**43.35**	1214	**114.52**
C	21	**14.68**	594	**56.03**
D	977	**683.21**	3582	**337.92**
E	46	**32.16**	1813	**171.03**
F	459	**320.97**	1273	**120.09**
G	229	**160.13**	2939	**277.26**
H	180	**125.87**	953	**89.90**
I	14	**9.79**	601	**56.69**
J	3	**2.09**	144	**13.58**
K	25	**17.48**	3388	**319.62**
L	11	**7.69**	1062	**100.18**

Table 7
Raw frequencies and normalised frequencies of emergency-related patterns in the aggregated corpus and in the aggregated corpus + RT

Facing #emergencies. The linguistic role of keywords, hashtags and retweets in communicating critical events

307

7. Final remarks

This paper aims at providing empirical evidence of the usefulness of digital and online devices, with particular emphasis on Social Media, in providing and spreading information to face and manage critical situations such as emergencies of different nature. By retrieving information from Twitter, a popular Social Media platform, four different corpora were created in order to find sub-categories of keywords that relate to events that shocked the public (and virtual) opinion, which were used actively to provide information for effective actions, also by means of tool-specific devices such as in the case of hashtags. An aggregated analysis of the four corpora provides evidence of the presence of common patterns in communicating crisis-related information. At the same time, the aggregated corpus shows a striking difference in terms of size and use of tool-specific device such as RTs in magnifying and amplifying the usefulness of information selected by the whole community, for the whole community. This difference can be noticed from different perspectives, such as corpus size; corpora including RTs are five to eight times larger than those with no retweets in terms of number of tweets, and five to almost ten times larger in terms of tokens. Corpora show a remarkable difference in terms of frequent pattern of use: in many cases, recurrent grammar and lexical patterns show exponential increase due to the use of Retweets, the latter guaranteeing the spread of a given message (and the resulting language patterns) which is observed from the comparison of the different corpora.

By identifying some specific language and channel-dependent variables, a discourse community on Social Media can be identified. Quality of information, especially the one generated and managed by institutional or verified accounts, proves to be reliable thanks to the mechanism offered by tool-specific devices such as RTs; finally, a wise use of other features of this kind, such as effective hashtags, is used to isolate the discourse within a specific field and purpose (Cunha *et al.* 2011).

Further studies of this nature may focus on other variables such as geo-location or stakeholders in the online community (i.e., numerical incidence of institutional vs. private accounts), or focusing on other Social Media platforms such as those commonly used by a global digital audience (such as Facebook) also for comparing the quality of information shared in emergency situations. Once again, it is important to underline that for the purpose of this study, Twitter has been preferred due to some favourable conditions imposed by its features: indeed, almost all data can be browsed and retrieved publicly. Furthermore, the typical use of language on this platform favours quick and prompt information that becomes effective thanks to the use of tool-specific devices such as RTs and hashtags. Evidence from this study shows that the online sphere and its growing community could use language for effective actions in the 'real' world. As a matter of fact, the kind of information provided via Twitter in crisis situation has real repercussions in terms of its performativity, and this is particularly true to divulge information that requires quick dissemination in order to manage emergency situations. In particular, institutional accounts (such as authorities or media outlets) are the early originators of messages which provide examples of language acts that become a standard to be followed during such situations and, by means of RTs, they prove to be statistically relevant due to their frequency. Emergencies, then, have a partner in its operative management: the solution of critical problems could be just one click away.

Bionote: Francesco Meledandri holds a PhD in Translation Theory and Praxis: European Intercultural Studies awarded by University of Bari in 2011. After a Research fellowship granted in 2013, he has been a lecturer (English language and translation) for eight years at University of Bari, Dept. of Humanities. He has been a lecturer in ICT solutions for Translation for six years, and a lecturer of IT (advanced course) for three years at the same University. Current research areas include: Translation Studies; Translation Theory and Praxis; CAT Tools; Machine Translation; Language of sports; Language of (and via) social media.

Author's address: francesco.meledandri@uniba.it

Facing #emergencies. The linguistic role of keywords, hashtags and retweets in communicating critical events

309

References

Agichtein E., Castillo C., Donato D., Gionis A. and Mishne G. 2008, *Finding High-Quality Content in Social Media*, in Najork M., Broder A., Chakrabarti S. (eds.), *WSDM '08 proceedings of the 2008 International Conference on Web Search and Data Mining*, Palo Alto, California, USA – February 11-12, 2008, ACM Press, pp. 183-194.

Baker P. 2006, *Using Corpora in Discourse Analysis*, Continuum, London/New York.

Baron N.S. 1984, *Computer Mediated Communication as a Force in Language Change*, in "Visible Language" XVIII [2], Spring 1984, pp. 118-141.

Barton D. and Lee C. 2013, *Language Online – Investigating Digital Texts and Practices*, Routledge, London/New York 2013.

Boyd D., Golder S. and Lotan G. 2010, *Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter*, in Sprague R.H. (ed.), Proceedings of the 43rd Hawaii International Conference on Systems Science (HICSS-43 2010), 5-8 January 2010, Koloa, Kauai, HI, USA, IEEE Computer Society, Los Alamitos, California (USA).

Bredl K., Hünninger J. and Jensen J.L. (eds.) 2014, *Methods for Analysing Social Media*, Routledge, London/New York.

Brooker P., Barnett J. and Cribbin T. 2016, *Doing social media analytics,* in "Big Data & Society" 3 [2], pp. 1-12.

Bruns A. and Burgess J. 2011, *The use of Twitter Hashtags in the Formation of Ad Hoc Publics*, paper presented at the 6th European Consortium for Political Research General Conference, University of Iceland, Reykjavik, 25-27 August 2011.

Bruns A. and Liang Y.E. 2012, *Tools and methods for capturing Twitter data during natural disasters*, in "First Monday" 17 [4]. http://firstmonday.org/article/view/3937/3193 (11.12.2020).

Bruns A. and Stieglitz S. 2012, *Quantitative Approaches to Comparing Communication Patterns on Twitter,* in "Journal of Technology in Human Services" 30 [3-4], pp. 160-185.

Burnap P., Williams M.L., Sloan. L., Rana O., Housley W., Edwards A., Knight V., Procter R. and Voss A. 2014, *Tweeting the terror : modelling the social media reaction to the Woolwich terrorist attack,* in "Social Network Analysis and Mining", 4 [1], p. 206. https://link.springer.com/article/10.1007/s13278-014-0206-4 (11.12.2020).

Crystal D. 2011, *Internet Linguistics: A Student Guide*, Routledge, London/New York.

Cunha E., Magno G., Comarela G., Almeida V., Gonçalves M.A. and Benvenuto F. 2011, *Analyzing the Dynamic Evolution of Hashtags on Twitter: a Language-Based Approach*, in Nagarajan M., Gamon M. (eds.), *Proceedings of the Workshop on Language in Social Media (LSM 2011),* Portland, Oregon 23 June 2011, Omnipress Inc., Madison WI (USA), pp. 58-65.

Dickinson P. 2013, *'B/w U & me': The Functions of Formulaic Language in Interactional Discourse on Twitter*, in "Linguistic Journal" 7 [1], pp. 7-38.

Giltrow J. and Stein D. 2009, *Genres in the Internet – Innovation, evolution, and genre theory*, in Giltrow J., Stein, D. (eds.), *Genres in the Internet – Issues in the theory of genre*, Benjamins, Amsterdam/Philadelphia 2009, pp. 1-26.

Herring S.C. 2001, *Computer-Mediated Discourse*, in Tannen D., Hamilton H.E. and Schriffin D. (eds.), *The Handbook of Discourse Analysis,* Blackwell, Malden/Oxford, pp. 612-634.

Herring S.C. 2008, *Virtual community*, in Given L.M. (ed.), *Encyclopedia of Qualitative Research Methods*, SAGE Publications, London, pp. 920-921.

Herring S.C. 2015, *Language and the Internet*, in Donsbach W. (ed.), *The Concise Encyclopedia of Communication,* Wiley Blackwell, Chichester (United Kingdom), pp. 322-323.

Herring S.C. and Androutsopoulos J. 2015, *Computer-Mediated Discourse 2.0*, in Tannen D., Hamilton H.E., Schriffin D. (eds.), *The Handbook of Discourse Analysis, 2nd ed.,* Wiley, Hoboken 2015, pp. 127-151.

Higham J.P. and Hebets E.A. 2013, *An introduction to multimodal communication*, in "Behavioural Ecology and Sociobiology" 67 [9], pp. 1381-1388.

Hoffmann C.R. and Bublitz W. (eds.) 2017, *Pragmatics of Social Media*, De Gruyter, Berlin.

Hughes A.L. and Palen L. 2009, *Twitter Adoption and Use in Mass Convergence and Emergency Events*, in Landgren J., Jul S. (eds.), *Proceedings of the 6th International ISCRAM Conference* – Gothenburg, Sweden, May 2009.

Jikoun V., de Rijke M., Ackermans P., Geleijnse G. and Weerkamp W. 2010, *Mining User Experiences from Online Forums: an Exploration*, in Hachey B. and Osborne M., *Proceedings of the NAACL HLT 2010*

Workshop on Computational Linguistics in a World of Social Media – Los Angeles, California, June 2010, Association for Computational Linguistics (ACL), Stroudsburg, PA (USA), pp. 17-18.

Johnson D. 2014, *Reading, Writing, and Literacy 2.0. Teaching with Online Texts, Tools, and Resources, K-8*, Teachers College Press, New York/London.

Jovanovic D., van Leeuwen T. 2018, *Multimodal Dialogue on social media,* in "Social Semiotics" 28 [5], pp. 683-699.

Kaplan A.M. and Haenlein M. 2010, *Users of the world, unite! The challenges and opportunities of Social Media*, in "Business Horizons" 53, pp. 59-68.

Kehoe A. and Gee M. 2011, *Social tagging: A new perspective on textual 'aboutness',* in Rayson P., Hoffmann S., Leech G. (eds.), *Studies in Variation, Contacts and Change in English – Methodological and Historical Dimensions of Corpus Linguistics*, Volume 6. http://www.helsinki.fi/varieng/series/volumes/06/kehoe_gee/ (11.12.2020).

Kern R. and Develotte C. (eds.) 2018, *Screens and Scenes: Online Multimodal Communication and Intercultural Encounters. Theoretical and Educational Perspectives*, Routledge, London/New York.

Kim D. and Vorobel O. 2017, *Discourse Communities: From Origins to Social Media*, in Wortham S., Kim D., May S. (eds.), *Discourse and Education. Encyclopedia of Language and Education*, Springer International Publishing AG, Cham, Switzerland, pp. 267-281.

Khosravinik M. 2017, *Social Media Critical Discourse Studies (SM – CDS)*, in Flowerdew J., Richardson J. (eds.), *Handbook of Critical Discourse Analysis*, Routledge, London, pp. 582-596.

Kress G., van Leeuwen T. 2006, *Reading Images. The Grammar of Visual Design*, Routledge, London and New York.

Leppänen S., Westinen E. and Kytölä S. 2017, *Social Media Discourse, (Dis)identifications and Diversities*, Routledge, London/New York.

Martin J.R. 2009, *Realisation, Instantiation and Individuation: some Thoughts on Identity in Youth Justice Conferencing*, in "D.E.L.T.A." 25, pp. 549-583.

Marwick A.E. and boyd d. 2010, *I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience*, in "New Media & Society" 13 [1], pp. 114-133.

McEnery T. and Hardie A. 2012, *Corpus Linguistics: Method, Theory and Practice,* Cambridge University Press, Cambridge/New York.

Mendoza M., Poblete B. and Castillo C. 2010, *Twitter Under Crisis: Can we trust what we RT?*, 1st Workshop on Social Media Analytics (SOMA '10), Washington, DC, pp. 71-79.

O'Reilly T. and Milstein S. 2012, *The Twitter Book (2nd ed.),* O'Reilly Media Inc., Sebastopol (California).

Olteanu A., Vieweg S. and Castillo C. 2015, *What to Expect When the Unexpected Happens: Social Media Communications Across Crises*, CSCW '15 Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, Vancouver, Canada, pp. 994-1009.

Partan S.R. 2013, *Ten unanswered questions in multimodal communication*, in "Behavioural Ecology and Sociobiology" 67 [9], pp. 1523-1539.

Reiß K. and Vermeer H. J. 2014, *Towards a General Theory of Translational Action. Skopos Theory explained*, Routledge, London/New York.

Shohamy E. 2006, *Language Policy – Hidden agendas and new approaches*, Routledge, London/New York.

Swales J. 1990, *Genre Analysis: English in academic and research settings,* Cambridge University Press, Cambridge.

Swales J. 2011, *The Concept of Discourse Community,* in Wardle E. and Downs D. (eds.), *Writing about writing,* Bedford St. Martins, Boston, p. 466-478.

Thurlow C., Lengel L. and Tomic A. 2004, *Computer Mediated Communication – Social Interaction and the Internet*, Sage Publications, London.

Zappavigna M. 2012, *Discourse of Twitter and Social Media: How We Use Language to Create Affiliation on the Web*, Continuum International Publishing Group, London/New York.

Zappavigna M. 2014, *Searchable audiences: The linguistic functions of hashtags*, paper presented at Australasian Audience Research Symposium: Data, context and meaning. University of New South Wales, 22 April 2014.

Zappavigna M. 2015, *Searchable talk: the linguistic functions of hashtags,* in "Social Semiotics" 25 [3], p. 274-291.