

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of Pragmatics

journal homepage: [www.elsevier.com/locate/pragma](http://www.elsevier.com/locate/pragma)

## Opening up #jesuisCharlie anatomy of a Twitter discussion with mixed methods



Marjut Johansson<sup>a, \*</sup>, Aki-Juhani Kyröläinen<sup>b, f</sup>, Filip Ginter<sup>c</sup>, Lotta Lehti<sup>d</sup>,  
Attila Krizsán<sup>e</sup>, Veronika Laippala<sup>d</sup>

<sup>a</sup> Department of French Studies, School of Languages and Translation Studies, University of Turku, FI-20014, Finland

<sup>b</sup> Department of Linguistics and Languages, McMaster University, L8S 4M2, Canada

<sup>c</sup> Department of Future Technologies, University of Turku, FI-20014, Finland

<sup>d</sup> School of Languages and Translation Studies, University of Turku, FI-20014, Finland

<sup>e</sup> Department of English, School of Languages and Translation Studies, University of Turku, FI-20014, Finland

<sup>f</sup> Department of Applied Linguistics, Brock University, L2S 3A1, Canada

### ARTICLE INFO

#### Article history:

Received 18 August 2017

Received in revised form 21 December 2017

Accepted 12 March 2018

### 1. Introduction

The terrorist attack on the editorial office of the left-wing satirical magazine, *Charlie Hebdo*, on January 7, 2015, was a prominent news event that was followed by global participation in social media. *Charlie Hebdo* has been a controversial weekly publication since its inception in the 1960s. Mostly using cartoons, it targets sensitive socio-cultural topics and aims at severe criticism of politics, culture, and religion. This has led to many controversies and court cases (cf. [Boumédiène et al., 2017](#); [Nugier and Guimond, 2016](#)). The most serious one was the republication of the Muhammed cartoons in 2006, which had originally been published by the Danish *Jyllands Posten* a year earlier. Although the republication led to heated debates in the French media, the court ruled that the limits of freedom of speech were not exceeded ([Nugier and Guimond, 2016](#): 46). In all, this incident had a strong impact on the Muslim community in France and led to a negative attitude against *Charlie Hebdo* ([Nugier and Guimond, 2016](#): 46).

The terrorist attack began with shootings at an editorial meeting, in which the two main perpetrators killed twelve persons. The next day, five more persons were killed in follow-up attacks. These events were brought to an end by the police, who killed the three terrorists. Investigation later proved that the attack had been domestic, as the terrorists were all French citizens ([Nugier and Guimond, 2016](#): 45). This disruptive news event was immediately given global media coverage and solidarity marches took place all over the world. Very soon after the attack had begun, the micro-blogging site Twitter exploded with tweets containing the hashtag #jesuisCharlie (*I am Charlie* in French), referring to the title of the magazine. The hashtag was formed on the basis of the logo created by art director Jacques Roncin. This was one of the most tweeted news events of its time, at the rate of 6500 tweets per minute ([Whitehead, 2015](#)).

\* Corresponding author.

E-mail addresses: [marjut.johansson@utu.fi](mailto:marjut.johansson@utu.fi) (M. Johansson), [akkyro@gmail.com](mailto:akkyro@gmail.com) (A.-J. Kyröläinen), [filip.ginter@utu.fi](mailto:filip.ginter@utu.fi) (F. Ginter), [lotta.lehti@utu.fi](mailto:lotta.lehti@utu.fi) (L. Lehti), [attila.krizsan@utu.fi](mailto:attila.krizsan@utu.fi) (A. Krizsán), [veronika.laippala@utu.fi](mailto:veronika.laippala@utu.fi) (V. Laippala).

<https://doi.org/10.1016/j.pragma.2018.03.007>

0378-2166/© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

This news event has been the object of various studies. In media sciences, the news stories on the *Charlie Hebdo* attacks unfolded over several days as the events took place and were characterized as a hybrid media event (Sumiala et al., 2016). Smyrnaio and Ratinaud (2017) conducted a network and cluster analysis identifying the communities that participated in the discussion and their topics. They discovered that the most important themes were a news event, freedom of speech, journalists, sympathy and condolences, expressions of horror and resentment, religious fanaticism, and judgment of the terrorists (Smyrnaio and Ratinaud, 2017: 7–8). Similar themes were found in linguistically-oriented studies of cartoons posted on social media (Bouko et al., 2017). Giaxoglou (2017) examined the hashtags #Charlie Hebdo and #jesuisCharlie as well as other hashtags used in sharing stories in the ecstatic participation in Twitter during this event. In addition, different, opposing counter-discourses emerged at this time, for instance in the forum discussions about the terrorist attack: users condemned the assault, but did not accept the satirical views of the *Charlie Hebdo* magazine (Badouard, 2016).

Our study differs from the previously mentioned studies in two fundamental ways: first, our study considers this massive and global social media discussion as a linguistic object. Our aim is to open up the large-scale data set in a way that will allow us to consider into the type of claims that have been made in this discussion and establish what we call here its anatomy. Second, we use a mixed methods approach that differs from the studies mentioned above to shed more light on tweeting after a disruptive situation. Our study opts to describe a type of Twitter discussion that does not represent what Zappavigna (2011) calls “ambient affiliation” in which participants share bonds and belonging nor it represents a Twitter discussion that Page (2012) calls a marketplace for self-branding and micro-celebrity. We start with the view that the *Charlie Hebdo* event was a hybrid news event that took place over days in several places and involved multiples actors (Chadwick, 2013; Sumiala et al., 2016: 100).

Our objective is to study the kind of social interaction that took place in tweets that were posted with the hashtag #jesuisCharlie. We focus on the kind of opinions that users tweeted by investigating the following questions:

- 1) What are the main topics in this Twitter discussion?
- 2) How are opinions expressed in typical tweets?

Our analysis aims at dissecting the nature of this #jesuisCharlie discussion with mixed methods. The main reason for taking this approach is the size of our data: we will focus on over 0.25 million tweets, which is why it will be necessary to use language technology methods to sort the topics in these tweets. Specifically, we will use clustering, a commonly used method in data analysis to identify previously unknown groupings in data (see Kaufman and Rousseeuw, 1990). In the second phase of the study, we will conduct a quantitative analysis that focuses on modality (Halliday and Matthiessen, 2014), and then perform a qualitative digital discourse analysis of topical clusters, concentrating on how opinions are expressed in the tweets, through the study of stancetaking and positioning (Davies and Harré, 1990; DuBois, 2007; Weizman, 2008).

Our first hypothesis about the topics of this Twitter discussion is based on the concept of the *topic of discourse* as “a property of the global meaning” (van Dijk, 1980: 41). This concept relates to the overall idea of what a discourse is all about, and a discourse topic also controls many other aspects of discourse. Regarding online discussions, Lambiase (2010) studied how topics evolved and shifted in a discussion list after a tragic news event in a small-scale corpus. She found that the global topic did not hold as the discussion progressed, but instead topic shifts and derailments were typical (Lambiase, 2010: 16). In contrast to Lambiase, we hypothesize that this online discussion does not consist of one topic, but contains multiple topics that are similar to the discourses that circulated cross-media (Smyrnaio and Ratinaud, 2017; Sumiala et al., 2016: 98).

Our second hypothesis focuses on how users express their opinions in their evaluation and their positioning about different topics. According to various studies, users express opinions that are affective, diversified, and polarized in several types of massive social media discussions. However, those studies are based on small qualitative data (Bou-Franch and Blitvich, 2014; Badouard, 2016), their approach is centered on media studies from theoretical (Marwick and Boyd, 2011), or network analysis perspective (Papacharissi, 2015). They have showed that during tragic news events, networked publics engage and show compassion with an affective response (cf. Papacharissi, 2015: 7). When sociopolitical issues arise in which values are concerned, diversification and polarization will emerge in the positioning that users take (Bou-Franch and Blitvich, 2014; Badouard, 2016; Smyrnaio and Ratinaud, 2017). In his qualitative analysis, Mercier (2016: 139) considers Twitter to be a space for counterpublics to formulate polemic and criticism and to destabilize opponents, especially in the political context. In a study that focused on discussion forum comments, Ferencik (2017) found that identification and expressions belonging to a moral order were prominent in counter-discourse. We hypothesize that stancetaking will be affective, diversified, and polarized in our massive discussion, but in contrast to above mentioned studies, our linguistically-based approach aims at explaining the types of affectivity, polarization, and diversification in different topical clusters within the global discussion.

Our paper is structured as follows. In Section 2, we discuss the functions of hashtags and tweets and we focus on Twitter communication in breaking global news events. In Section 3, we discuss stancetaking, especially from the perspective of modality, as well as positioning as characteristics of opinions in tweets. Then, in Section 4, we present our data and methodology. Our results on the topics found in the data are described in Section 5; Section 6 discusses the quantitative patterns on modality, and Section 7 contains a digital discourse analysis of the opinions. Section 8 concludes the paper.

## 2. Tweets, hashtags, and Twitter discussions in global news events

The functions of tweets, or micro-posts, vary according to several factors (Page, 2012: 183). These factors are linked to the type of event or content that the users are commenting on and the topics that emerge from them. Tweets can have several discursive functions and can index a positioning (Page, 2012: 184; Zappavigna, 2014: 211; Scott, 2015). One of the most characteristic elements of the tweet is the hashtag, which allows for users to express what the tweet is about and to index it as searchable talk (Zappavigna, 2011). As mentioned, the data collection of the present study is based on one specific hashtag, #jesuisCharlie. In her study, Giaxoglou (2017) calls it a metanarrative hashtag.

The hashtag, that is, the combination of the “#” sign and a word or phrase, is a metadata tag originally introduced to enable the search of tweets including a given hashtag. As Scott (2015) argues, the functions of hashtags today are not restricted to search. Scott (2015) investigates hashtags from the perspective of relevance and claims that the main function of hashtags is to guide readers' interpretations, that is, to make contextual assumptions accessible to a wide range of readers. Zappavigna (2011) explores the search function enabled by hashtags but combines it with the evaluative function: she views hashtags as a means of creating communities of shared value since they enable affiliation via findability (Zappavigna, 2011: 789). Puschmann (2015) also highlights the community building function of hashtags, in addition to the playfulness dimension of hashtag use. Puschmann (2015: 30) views the communities that emerge from shared hashtag use as “groups of users that form in an ad-hoc fashion around a particular hashtag and have a relatively loose social structure” (see also Bruns and Burgess, 2011). Page (2012: 196), however, considers that hashtag use in micro-posts involves participants talking *about* the same topic rather than talking *with* each other. In Page's (2012) view, hashtags are a resource of self-branding realized in two main types of hashtag: topic-based (i.e., making the topic of a tweet visible) and evaluative (i.e., emphasizing stance). The #jesuisCharlie discussion contains both topic-based and evaluative hashtag use (see Section 2) as examples (1) and (2) illustrate<sup>1</sup>:

(1) Topic-based hashtag use:

#JeSuisCharlie March in Paris featured leaders of Europe who will throw your butt in jail for “Hate Speech”

(2) Evaluative hashtag use:

I don't always cry when I watch the news, but when I do it's because it's the crack of dawn in #paris #JeSuisCharlie

In both examples, the hashtags are both topic-based and evaluative. However, it appears the two types of hashtag use are emphasized differently. In example (1), the hashtag is a modifier in the subject of the clause and part of the theme, instead of being the rheme. Even if the tweet contains also evaluation of the motives of the participants of the march, the main function of the hashtag is to make the topic of the tweet visible: it names the march. In contrast, example (2) illustrates how a hashtag can emphasize stance. The author of example (2) expresses sadness over the events and thus claims solidarity with the victims or at least with others sharing these emotions. In this case, the hashtag #jeSuisCharlie can be considered evaluative because it reinforces the stance (emotion, solidarity), in addition to indicating the topic of the tweet.

Twitter is a space of “mass, real-time expression of opinion and sentiment” about public- and private-sphere events and matters (Zappavigna, 2014: 209–210). As a platform, it allows the networked public to emerge easily, but it belongs to the digital culture of complex social positioning and relationships. It is not surprising, therefore, that several studies have focused on geolocation and networks of users (Smyrniaios and Ratinaud, 2017). Twitter discussions have emerged on several occasions relating to global news events and have been referred to as “Twitter revolutions,” such as during the Arab Spring movement (Bruns et al., 2013). This multilingual discussion has been studied through the use of the #egypt and #libya hashtags, and researchers have found that most of the discussion was conducted by those who were interested in the Egyptian case (Bruns et al., 2013). In the case of the terrorist attack in Norway in 2011, the main discourses were centered on the nation, solidarity, and the attacks (Eriksson, 2016). According to Eriksson (2016: 366–368), the tweets represented the post-trauma discourse of coming together, through which the public could share collective self-reflection after a tragic event. In counter-discourse, tweets are also emotional, but they may have different targets. They can oppose the media and try to destabilize its information or seek support for the author's own cause (Mercier, 2016: 141), as in the case of activism.

Yang (2016) describes the case of hashtag activism through the example of #BlackLivesMatter: how it was adopted, how its spread created an alternative narrative to that presented in the professional media, and how it eventually turned into a social movement for racial justice. The #BringBackOurGirls campaign is another well-known example of the widespread use of a hashtag as political contention, that is, of hashtag activism. This hashtag was used to appeal for the release of over 200 girls who had been held as hostages by the Boko Haram terrorist organization in Nigeria since April 2014. However, it is also used for the more general fight of securing the right to education for all girls (see Chiluwa and Ifukor, 2015). The failure of many hashtag campaigns to lead to concrete, measurable results in a reasonable time span have caused them to be viewed in a negative light. For instance, Morozov (2011: 189–190) describes the blaming of users for being “slacktivists,” that is, for being too slack to “really act” and just spreading a hashtag. A slightly less derogatory term is “clicktivism,” which signifies the minimal effort of clicking in social media as activism, connoting inefficiency.

<sup>1</sup> We use all tweets as they are and do not correct grammatical or lexical errors. We use italics to refer the magazine *Charlie Hebdo* and brackets [ ] in order to explain the contents of the tweets when necessary. We use the orthography #jesuisCharlie for the basic hashtag, but, in the tweets, the orthography of these can vary. To highlight linguistic elements, we use underlining and italics in the examples.

In the next section, we turn to the question of how opinions expressed in tweets can be analyzed. We will do this from three perspectives: modality in utterances, linguistic marking of evaluation in stancetaking, and communicative actions of discursive positioning.

### 3. Opinions in tweets: evaluating, stancetaking, and positioning

On Twitter, users express their opinions in micro-posts. Linguistically, opinions are evaluative comments by which social actors engage in a communicative activity that considers an event or an issue, that is, an object of talk (cf. DuBois, 2007; Jaffe, 2009; Thompson and Hunston, 2003). Here, we will consider opinions from the perspective of stance and positioning.

In linguistic literature, *stance* and *positioning* are heterogeneous concepts that are used in different scholarly traditions (see Thompson and Hunston, 2003; DuBois, 2007; Jaffe, 2009). The analytic traditions are numerous: corpus linguistics, socio-linguistics, systemic-functional linguistics, critical discourse analysis, and interactional approaches, just to mention a few (Jaffe, 2009: 2). As a linguistic concept, stance has been approached as a grammatical and lexical stylistic device for marking attitudes, feelings, and judgments (Biber and Finegan, 1989) and for evaluation of authorial stance (Hunston and Thompson, 2003), attitude (Halliday, 1985), and appraisal (Martin and White, 2005). We employ two levels of study, the first from the utterance-internal perspective of modality and the second of stance as a communicative act.

First, modality can be considered as a dimension of stance. *Modality* refers to how the propositional content is modified by a speaker (Halliday, 1985, 2014). Modality expresses the ways interactants relate to their propositions in terms of validity, that is, *epistemic modality*, or indicate their confidence in the success of exchange, that is, *deontic modality* (cf. e.g., Halliday and Matthiessen, 2014: 691; Kaufmann et al., 2006; Fintel 2006; Kiefer, 1987). Halliday and Matthiessen (ibid.) maintain that epistemic modality or modalization can be divided into categories of probability or usuality, while deontic modality or modulation refers to either obligation or inclination. Probability stands for the likelihood of an event or a state of being expressed in the proposition to be true, ranging from possible to certain. Usuality then indicates the frequency of how often a proposition is valid, ranging from sometimes to always. Obligation is related to the degrees of state of affairs in terms of the content of the proposition, and it ranges from being allowed to being required. And, finally, inclination signals levels of willingness, related to what should be the state of affairs, ranging from being willing to being determined. Modality, therefore, can be seen as communicating one's stance on the "truthness" of one's propositions, and, at the same time, it can also display the amount of responsibility taken for the point of view being expressed in a proposition.

Second, we consider *stance* as a linguistic and a social act (DuBois, 2007: 141). When taking a stance, social actors evaluate objects, and when *positioning*, they express their subjectivity and establish intersubjective relationships with others (cf. DuBois, 2007: 158). This view has been proposed for spoken interaction, but we consider that it can be adopted in written interaction of tweets as well. In stancetaking, social actors may display an affective, epistemic, or any other stance that grounds them in situations giving them authentic voice and authority (see Jaffe, 2009: 8–9). In their reactions, they may communicate a private sphere, individual feeling, or they may take a stance on societal issues, showing their expert knowledge. Furthermore, by positioning, social subjects engage themselves in a "discursive process whereby selves are located in conversations as observably and subjectively coherent participants (...)" (Davies and Harré, 1990: 48). In other words, social actors inscribe themselves into situations and determine relationships with others (Weizman, 2008: 14). In positioning, social actors index social identities that are related to the social situations and contexts in which they act (Jaffe, 2009: 9). Furthermore, when positioning, social actors *align* or *disalign* with others (DuBois, 2007: 163). In the next section, we describe our data and the different phases of our methodology.

### 4. Data and mixed methods

The data consist of tweets with the hashtag #jesuisCharlie posted between January 7 and 14, 2015. The tweets were collected using the *YourTwrapperKeeper* application (Bruns and Liang, 2012)<sup>2</sup>. The initial data includes 1.2 million tweets in 51 different languages, and the most tweeted languages are French, English, and Spanish. For the analysis, we chose the tweets in English. This set contained approx. 0.5 million tweets, but after retweets were excluded, the data comprised 234,363 tweets, which is similar to the corpus used in Smyrniaios and Ratinaud's study (2017: 5).

The research design is sequential, based on mixed methodology consisting of three phases (cf. Creswell, 2013: 14–16). First we use a computational approach with language technology methods which treat data as decontextualized; second, a linguistic and grammatical approach is taken to the modality of utterances in topical clusters; and third, we conduct a qualitative digital discourse analysis based on the communicative action in the digital context.

In the first phase, to examine the distribution of different topics across the tweets, we applied several language technology methods. First, the tweets were part-of-speech tagged with the twitter tagger, GATE (Derczynski et al., 2013), and then the data were cleaned from function words such as *of*, *it* and *like*, that are very frequent but carry little meaning and thus corrupt the analysis (see Jurafsky and Martin, 2017: 143, 164). This is a very standard technique in natural language processing to ensure high quality results. Benefiting from the part-of-speech tags, we included in the analysis only words that were tagged as verbs or nouns. Tweets that did not include any words after this cleaning process were deleted from the analysis.

<sup>2</sup> The data was collected by Marco T. Bastos and Raquel Recuero.

Second, to carry out a quantitative analysis of the semantic information and the topics expressed in the tweets, we constructed a semantic vector representation for each tweet using a neural network model, word2vec (Mikolov et al., 2013). Word2vec models learn to identify semantically similar words based on a large set of training data. The theoretical foundation of the method is in the distributional hypothesis, which considers that similar words are used in similar contexts (see Firth, 1957; Harris, 1968; also Gries, 2012). By the examples provided in the training data, word2vec learns to predict similar vectors for words occurring in similar contexts, which thus are also semantically similar. These vectors allow for the quantitative analysis of the word semantics. In this experiment, the word2vec model was trained on the ukWaC corpus (Baroni et al., 2009). To get vectors for the tweets, we took the averages of the tweet word vectors.

Third, to group semantically similar tweets together, we clustered the semantic vectors of the tweets in R (R Core Team, 2015), using k-means clustering (Skmeans package, Hornik et al., 2012). Clustering is a standard method in information retrieval to group similar instances, such as similar tweets, to clusters in previously unseen data (Kaufman and Rousseeuw, 1990). The result of this clustering into five main topical clusters is described in Section 5. Finally, we estimated the keywords of each cluster using tf-idf (Jones, 1972). Tf-idf, or *term frequency–inverse document frequency*, is a frequent measure to estimate keywords to documents by analyzing the importance of words in a collection of documents. By taking into account the number of documents in which the word is found, tf-idf gives less importance to very frequent words which typically are not very informative about the text topic (Manning et al., 2008).

In the second phase of investigation, we examined the most typical tweets in the five clusters. These tweets were selected on the basis of the topical clustering of the data. From each cluster, 40 the most typical tweets were selected, except for Cluster A, where 26 tweets were collected after discarding repetition in the data of the 40 most typical tweets. (For further explanation of the topical clustering, see Section 5.) The analysis focused on modality and on quantitative patterns in the usage of modal operators—that is, modal auxiliaries and adverbs in each cluster—in order to compare the frequencies of modalized typical tweets across the clusters. The goal was to determine whether there are differences in the ways the propositions are made in the typical tweets for each topical cluster in terms of modality. That is, we were interested in whether tweeters vary their stance on the truthness of their propositions and the amount of responsibility across the topical clusters. The data was annotated manually with the help of the UAM CorpusTool program. The analytic categories used were based on the four categories of deontic and epistemic modality (Halliday and Matthiessen, 2014), that is, probability and usuality (categories of epistemic modality or modalization) and inclination and obligation (categories of deontic modality or modulation) and their values (cf. Section 3). The result of the quantitative patterns is presented in Section 6.

The third phase consists of qualitative analysis based on digital discourse analysis, which is based on discourse analysis, sociolinguistics, and sociopragmatics. Here, we analyze how users express stances and what is their discursive positioning in the most typical tweets of each cluster (Section 7). In each tweet, we identify objects of stance and the function of hashtags (cf. Sections 2 and 3). We analyze the explicit or implicit linguistic means, communicative acts and activities by which the evaluation takes place, paying attention to the subjectivity, indexicality, and affectivity in tweets (cf. Section 3).

## 5. Most typical topics in clusters

The clustering resulted in a solution with a total of nine clusters offering the best fit. Four of the clusters gather uninteresting material for this study, such as spam, and tweets without any text. These are excluded from further examination, and we focus on the five clusters with linguistic content. Table 1 below presents the results of the analysis in the first phase for these five clusters (labeled here as A, B, C, D, E), using selected keywords counted with tf-idf (Manning et al., 2008) and their general topics, revealed through manual analysis of the most typical tweets of each cluster. The most typical tweets were selected based on their distance to the estimated cluster center. These can be considered as the ones that best reflect the clustering principle, that is, in this study, the topics of the clusters.

**Table 1**  
Clusters in the #jesuisCharlie discussion.

	Selected keywords	Topics
Cluster A 17,845 tweets	Edition, newspaper, cover, paper, special	<ul style="list-style-type: none"> <li>• The new issue of Charlie Hebdo</li> <li>• Newspapers: Iranian newspaper banned, Fox sued</li> </ul>
Cluster B 16,572 tweets	Paris, San Francisco, London, Korea, solidarity, march, people	<ul style="list-style-type: none"> <li>• Politicians' reactions and actions</li> <li>• Marches for peace</li> </ul>
Cluster C 27,777 tweets	Think, speech, says, wonder, right, forget, know, want	<ul style="list-style-type: none"> <li>• Critical views against the satire, at the same time defenses of freedom of speech</li> </ul>
Cluster D 27,105 tweets	Speech, freedom, world, follow, expression, support, media	<ul style="list-style-type: none"> <li>• Direct insults at the supporters of jesuischarlie</li> <li>• Support for freedom of speech and Charlie Hebdo</li> <li>• Links to other similar events</li> </ul>
Cluster E 21,859 tweets	Muslims, terrorism, people, religion, Islamophobia	<ul style="list-style-type: none"> <li>• Jihadists and peaceful Muslims</li> <li>• Extremists</li> </ul>

A preliminary analysis of the clusters indicates the following main topics in the data: news issues and media (Cluster A), politicians' reactions and actions as well as marches for peace (Cluster B), critical views and defense of freedom of speech (Cluster C), support for freedom of speech and similar events (Cluster D), and Muslims and extremism (Cluster E). These are

similar to the clustering results obtained by Smyrniotis and Ratinaud (2017), but instead of 12 clusters obtained we obtained five clusters (See Introduction). In the next section, we examine each cluster in terms of quantitative patterns of modality.

## 6. Quantitative patterns of modality

The analysis concentrated on epistemic modality and its categories of probability and usuality as well as on deontic modality and its categories of obligation and inclination, as shown in the following examples:

- (3) I don't mind saying that cats are evil. It may [probability] be against your religion to say so, but it's not against mine. #JeSuisCharlie
- (4) The satire on Charlie Hedbo is crazy, but I will [inclination] defend freedom of speech, because when my time comes, someone will [inclination] too #JeSuisCharlie

Example (3) illustrates the way epistemic modality restricts the truthness of a proposition by reducing it to probability. The category of deontic modality in the data is illustrated by Example (4), which also demonstrates the way inclination contributes to reduced responsibility (in contrast to an unmodalized construction) and expresses personal involvement at the same time.

The findings of the contrastive analysis of the types of modality used across the clusters are summarized in Fig. 1. The percentage values were calculated with respect to the amount of typical tweets in each cluster (n = 40, except for Cluster A where n = 26; see Section 4 for explanation).

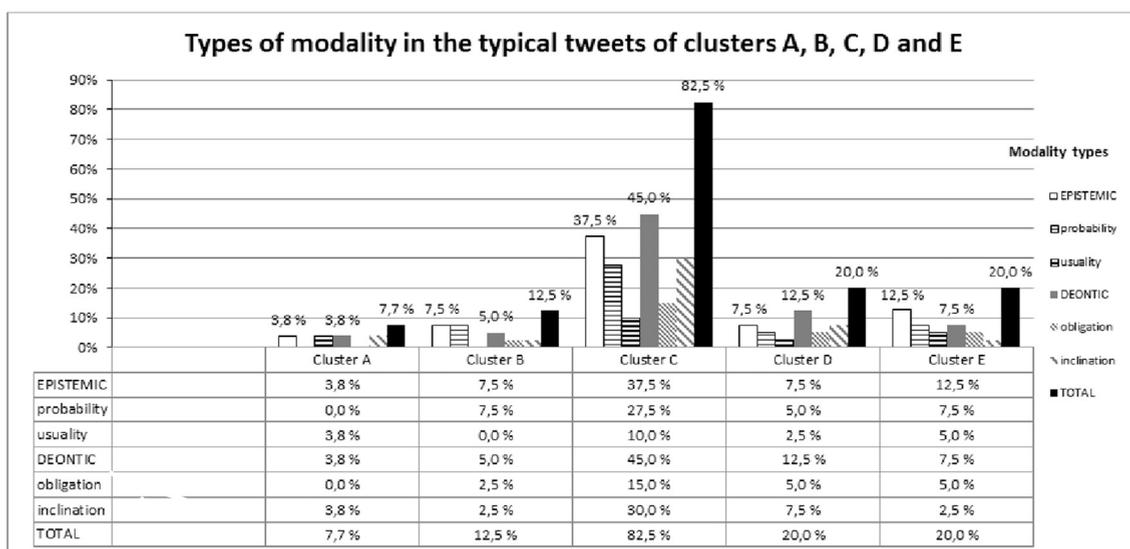


Fig. 1. Types of modality in the most typical tweets.

With regard to modality, the tweets in Cluster A are characterized by the lack of modality: there are only two cases, one inclination and one usuality. This implies that the typical tweets in this cluster may be less reduced in terms of the truthness of their propositions and their validity. However, these findings are inconclusive, as the difference between Cluster A and the rest of the clusters (except for Cluster C) for the total amount of modality cannot be considered statistically meaningful with certainty: the difference is not significant between Clusters A and B ( $p = 0.53526$ ) and Clusters A and D or Clusters A and E ( $p = 0.17384$ ). Cluster B contains the second lowest amount of modality among the five clusters. However, there is no statistically significant difference in terms of modality features between this cluster and the rest of the clusters with the exception of Cluster C.

In terms of the total amount of modality, the typical tweets of Cluster D resemble those of Cluster E (eight cases each). The types of modality used in the two clusters are somewhat different as there are more cases of deontic modality (five cases versus three) in Cluster D, while epistemic modality can be found more in Cluster E (five cases versus three). However, these differences are not statistically significant and neither are the differences with respect to total cases of modality between Cluster E and the rest of the clusters, except for Cluster C, which stands out in terms of modality features.

The frequency related findings for Cluster A indicate that critical views of the satire, defense of freedom of speech, and insults of supporters of #jesuisCharlie as the key topics of Cluster C seem to inspire a more interactive, argumentatively oriented, and personally involved stance than other topics, such as the new issue of *Charlie Hebdo* (Cluster A), politicians' reactions and actions (Cluster B), support for freedom of speech and links to similar events (Cluster D), and extremists vs.

peaceful Muslims (Cluster E). This is based on the statistically significant differences between Cluster C and the rest of the clusters in terms of 1) the total number of cases of modality in general, 2) the total number of cases of epistemic modality, 3) one of its subcategories: probability, 4) the total number of cases of deontic modality, and 5) one of its subcategories: inclination. These findings highlight that the typical tweets of Cluster C are more modalized than those in the rest of the clusters in terms of both validity and the tweeters' confidence in the success of the exchange. Since the dominance of Cluster C in modality structures is the most apparent in relation to probability (epistemic modality) and inclination (deontic modality), as shown by Examples (5) and (6) below this dominance also implies that typical tweets in Cluster C are more limited in terms of truthness and responsibility than the typical tweets of the rest of the clusters.

- (5) You know Tiffany, you have a lot of blood on your hands #JeSuisCharlie. This crazy world war probably wouldn't [probability] be going on if it wasn't 4 u.  
 (6) It's a great chance to attack islam through #JeSuisCharlie. I can [inclination] tell you that as a muslim I don't support those killings. I simply am

The quantitative dominance of cases of probability and inclination such as the ones found in Examples (5) and (6) above implies that, in Cluster C, typical tweets are more subjective than are those in the rest of the clusters, as they operate with more cautious claims, rather than simply stating facts in terms of their truth content. Furthermore, these claims are tied more closely to inclination; that is, while responsibility for the claims is less strongly expressed than it would be in unmodalized constructions, it is at the same time more strongly connected to personal motivation. In the next section, we will analyze stancetaking and positioning in tweets.

## 7. Stancetaking and positioning in the #jesuisCharlie Twitter discussion

In this section, we analyze the tweets in each cluster and their typical features, such as subjectivity and objectivity. We examine what are their objects of stance, how the (affective) evaluation is expressed and what function hashtags play in it. In addition, we pay attention to indexing time and place in stancetaking and positioning vis-à-vis this event.

### 7.1. Cluster A: News and solidarity

This first cluster contains tweets that take either the *news event and media* or *solidarity* as their object of stance. In the following example, the expression of solidarity is based on the topical use of the slogan #jesuisCharlie both at the beginning and at the end of the tweet:

- (7) Je suis Charlie = I am Charlie. It's a messg of solidarity w/the victims of the Attack on French Magazine Charlie Hebdo #JeSuisCharlie

In the beginning, this tweeter quotes the slogan and translates it into English. Through a metapragmatic utterance (*It's a messg of solidarity w/the victims*), the user expresses an explicit act of identification and solidarity (cf. [Giaxoglou, 2017](#)). The hashtag is an evaluative marker at the end of the tweet, reinforcing the message.

This cluster is characterized by its references to *Charlie Hebdo* as *French Magazine Charlie Hebdo* and its new issue published after the attack. In addition, there are several references to other media and news sites as well, such as *News Sentinel*, *Iranian Paper*, *Fox*, *#BBC*, *@YahooNews*, and *French News Website*, and a popular show (*The Simpsons*) is also mentioned. The cluster contains micro-posts that report of publishing and banning news in Iran. Celebrities' names figure in micro-posts that report about them expressing solidarity with the victims. In this respect, these tweets reflect the hybridity of this news event. Concretely, users index different temporal phases in the course of the news event:

- (8) Survivors' issue: Charlie Hebdo's new cover has Prophet holding 'Je Suis Charlie' sign: French ... <http://t.co/KHQcUghA4B> #JeSuisCharlie

In the micro-post (8) above, the user shares information about the *Je suis Charlie* slogan on the new cover of the *Charlie Hebdo* issue after the attack. The stance is not modalized, as the tweet conveys information. In the next Example (9), the user makes a demand:

- (9) Man up BBC and show the cover of the latest edition of Charlie Hebdo what happened to Je Suis Charlie! #bbcnews #JeSuisCharlie

The demand (*man up, show*) above is targeted at a media company that has not published the cover of the new *Charlie Hebdo*. In these Examples (8, 9), the #jesuisCharlie hashtag has the evaluative function in the end position of the tweets. In addition, it is possible to see how the event evolves, from the first moment of the event to later phases in these tweets. However, in some of the tweets, users tweet about the news but express opposite views:

- (10) “Je Suis Charlie” – but Not in My Backyard: Following the Charlie Hebdo attack, world leaders w ... <http://t.co/AmQxxdUAcw> #JeSuisCharlie

This tweet (10) contains a link to a news article with the headline, “‘Je Suis Charlie’ - but Not in My Backyard.” The article is published on a non-profit U.S. news website and reports that some of the world leaders who have participated in the marches and expressions of solidarity have themselves committed crimes. Therefore, this micro-post can be considered as a counter-discourse that opens a critical perspective on the emotions created by the traumatic event.

In sum, these users position themselves as critical readers of the news event and observe how the journalistic media talk about the event or whether the media has left something out. In the latter case, tweeters also demand that news be published about the course of the events. In addition, they express their solidarity and thus share the emotion that allows them to come together (cf. Eriksson, 2016). In the opposing case, they show that the expressed emotions are not sincere, thus forming a counter-discourse.

### 7.2. Cluster B: Expressions of solidarity through the translocal places

In this cluster, the objects of stance concern the marchers of solidarity and political reactions. On closer examination, the main evaluation in these tweets shows how the users index place in their tweets:

- (11) *Amazing scenes in paris, not just france, but the world coming together* in a fight against terror and for freedom of speech #JeSuisCharlie  
 (12) *Great to see this show of unity* #jesuischarlie - BBC News - France attacks: World leaders link arms to march in Paris <http://t.co/BxUwcT7iGe>  
 (13) Newcastle, Australia #French community comes together as millions march in the wake of #Paris attack #jesuischarlie <http://t.co/nswB47deaM>

In these examples, users index the places where they have participated in solidarity gatherings or where they have witnessed them. This marking of geolocation is manifested also in the use of hashtags in Cluster B. The tweets contain hashtags indicating place names, such as #France, #Paris, #Verviers, #nigeria, #irak, #pakistan, and #America. In these micro-posts, users' stances are affective: they align with others by using affective adjectives. In these Examples (11, 12, and 13), users exalt solidarity gatherings (*amazing, great*) and underline the sharedness of the experience (*world coming together, unity, French community coming together*). The links in the Examples (12, 13) lead to news articles that talk about the solidarity marches. These tweeters' positionings display an embodied presence in places witnessing marches or talk about how they have followed media stories about the marches. The type of expressions can be characterized as experienced spaces (Certeau, 1990). Interactants position themselves as public members who are politically and/or socially aware of the gravity of the situation and want to express feelings of community through translocal networking and expressions of affectivity.

In this cluster, however, some tweets express a counter-stance toward the sharedness of the experience. This type of stance is expressed in terms of the ignorance of events taking place or direct denial of solidarity:

- (14) Why have people protested against the Charlie Hebdo cartoons in London? It was a French magazine published in France. #JeSuisCharlie  
 (15) Where was the President of the USA? Why wasn't he there? World leaders unite for massive Paris march #JeSuisCharlie <http://t.co/R2VBjapboR>

It is difficult to interpret tweet (14): *Why have people protested against the Charlie Hebdo cartoons in London?* This question is either a rhetorical one or an information-seeking one. In the former case, the tweet expresses an ironical stance toward the expression of solidarity; in the latter, which is more probable, the user does not know what *Charlie Hebdo* is and the question is information seeking. In contrast to this bafflement, Example (15) conveys blame. The tweet contains two questions that can be interpreted as rhetorical questions that accuse the President of the U.S. (Obama, at that time) of not participating in the march of solidarity. Whereas there was a critical stance toward world leaders for participating in the solidarity marches in Example (10), here the link is to an article in the *Washington Post* that simply reports the marches. These two tweets show that the boundaries between clusters are fuzzy.

To resume, the cluster contains tweets that can be interpreted in two ways: those that express their enthusiasm about the solidarity through experienced spaces and thus index time and place. The other tweets are ironical or blaming. Users in this latter type of tweet position themselves against the solidarity or are critical about certain aspects of it, thus producing counter-discourse (cf. Badouard, 2016). Thus, this cluster contains tweets that are polarized.

### 7.3. Cluster C: Freedom of speech

In this cluster, the object of stance is often freedom of speech. As shown in Section 6, this cluster differs considerably from the other clusters as regards modality. In addition, Cluster C contained very few other hashtags than #jesuisCharlie. What is typical of these tweets is the use of the first person singular pronoun “I” with opinion and speech verbs and verbal

constructions (*I stand, I said, I don't give a shit*) as well as cognitive and epistemic verbs (*I understand, I didn't know*) (see Section 6). In this cluster, tweets are communicative acts in which the users either defend or express reservation about the expression of freedom of speech. These interactants take sides and align or disalign with others, as in the following examples:

- (16) *I stand* with France and as a New Yorker *I understand* how Paris is feeling right now and want you to know you are not alone. #JeSuisCharlie  
 (17) #JeSuisCharlie although *I didn't know* it existed before today, and it's mocking lots of the things *I value*, yet today *I'm* with #CharlieHebdo

In these Examples (16, 17), the users simultaneously express an identification with the expression of solidarity. In Example (16), the hashtag is in the evaluative position at the end of the micro-post. This user discloses their hometown, thus revealing a local identity. The micro-post contains a recognition of the emotion in France and an expression of support. In Example (17), the hashtag is evaluative in the beginning of the tweet, and another, shorter hashtag #CharlieHebdo is included in the end. The micro-post begins with a negative statement in which the user reveals that they have learned a moment ago about the magazine and its nature, yet they express support.

Despite the whole-hearted expression of support, this cluster still contained polarized tweets. Some of the tweets express a counter-stance, but several users refuse to take sides, and they express solidarity. These tweets display the opposition in violent terms:

- (18) @liamjlhill Yeah they can, like *I said I condemn* it, but saying #JeSuisCharlie is not the way to do it for me & many others  
 (19) *I am here to tell* French people that you are the lowest form of life on earth. *I don't give shit* about these. #JeSuisCharlie Charlie Hebdo2  
 (20) That's why *I couldn't get down* w that #JeSuisCharlie nonsense. You incited extremists, got a reaction and *now I should feel sorry for you?*

In Examples (18) and (19), the stance is similar to the kind [Badouard \(2016\)](#) found in his analysis: it does not accept the terrorism but it has reservations about the Twitter discussion or the magazine itself. Both these Examples (18, 19) contain speech act verbs (*say, tell*) and a verb and a verbal construction denouncing the support (*condemn, give a shit*). In some of the counter-stances, there were expressions of doubt (*Do you even know what the magazine is about?*) and a refusal. Example (20) is an extreme case of irony and bashing, as this interactant first condemns the Twitter discussion (*That's why I couldn't get down w that #JeSuisCharlie nonsense*), then makes an accusation (*You incited extremists, got a reaction*), and ends with an ironical rhetorical question (*now I should feel sorry for you?*).

In sum, positioning as regards freedom of speech is collective: either it is shared or users are against it. The counter-discourse contains more severe expressions than that in other clusters: it is ironic or contains extreme expressions, such as bashing.

#### 7.4. Cluster D: Other's reactions and group belonging

In this cluster, the object of stance is often reactions to what people are doing, other users, and group belonging. This cluster contains several other hashtags than the #jesuisCharlie hashtag, including hashtags of media and media events such as #SNL (*Saturday Night Live*), #TopGear, #bbcaq, and #GoldenGlobes. There are also opinions used as hashtags, such as #Opinion, #FreedomofSpeech.

In these tweets, users take up similar events and freedom of speech as in the previous Cluster C, but here they do not take a stand as explicitly. Instead, users evaluate expressed opinions. They signal how they are positioned toward what people have said:

- (21) *I could not be more proud* of how those around the world are handling the attack on freedom writers and illustrators in Paris. #JeSuisCharlie  
 (22) So is Charlie Hebdo still relevant or has everyone just [given] up and moved on? the fight for free speech isn't over, *folks*. #JeSuisCharlie  
 (23) *It might be controversial* but isn't featuring Muhammad on your front cover so soon after such a tragedy asking for trouble? #JeSuisCharlie

Either they express an affective stance of boosting evaluation (*I could not be more proud*) (Example 21) or they ask rhetorical questions (18, 23). Users are indirectly conveying their point of view, either urging the continuation of support for freedom of speech (22) or warning about the new situation (23). In their positionings, these users express inclusion with or exclusion from a group and its opinions. In addition, they evaluate the type of journalism of *Charlie Hebdo*.

#### 7.5. Cluster E: Religion and terrorism

In this last cluster, the objects of stance are mainly religion and terrorism. Among the five clusters analyzed, this one features the highest number of other hashtags than #jesuisCharlie. Most of the hashtags express an ideological or religious

issue in some way: #Muslims, #JeSuisAhmed, #WholsMuhammed, #Islam, #Peace, #Nemtsov, #Israel, #Jews, #ukip, #terrorism, and so on. Indeed, after immediate reactions to the event, the #JeSuisCharlie hashtag was transformed into different versions, some expressing criticism toward the magazine *Charlie Hebdo* and/or underlining the complexity of the events and their outcome. These include the hashtags #JeNeSuisPasCharlie (*I am not Charlie*), #nous sommes Charlie (*we are Charlie*), #jesuisAhmed (referring to the police officer killed in the course of events), and #jesuiskouachi (*I am Kouachi*). This latter hashtag refers to the last name of the two perpetrators and is one indicator of the counter-discourses during the event (cf. [Giaxoglou, 2017](#)).

In these micro-posts, users evaluate religion as in Examples (22) and (23):

- (24) Also, Islamic Jihadism put Muslims in *danger* more than it defends them. *War and violence is not in religion.*  
#JeSuisCharlie
- (25) #JeSuisCharlie #JeSuisAhmed. *Religion not problem; stupid is!* Muslims Around The World Condemn Charlie Hebdo Attack <http://t.co/QSf023Y6yT>

The hashtags are either evaluative (24) or topical (25) in these examples. In both these micro-posts (24, 25), users evaluate the consequences in a negative manner (*danger, problem*), signaling the risks of this attack to Muslims. Both examples contain slogan-like negative utterances in the middle of the tweets: *War and violence is not in religion* (24) and *Religion not problem; stupid is!* (25). Example (25) ends with a newspaper headline and a link to a *Huffpost* article that condemns the attack. A user also takes a stance against the extremist views expressed:

- (26) *It's sad to see* such large anti-Islam protests in Germany. Islamic Extremists should not be confused for Muslims  
#JeSuisCharlie #JeSuisAhmed

In Example (26), there are two evaluative hashtags in the end of the tweet, both of which are identity slogans for Charlie and Ahmed. The tweet begins with an expression of emotion (*it's sad to see*). The micro-post contains an explanation in the middle (*Islamic Extremists should not be confused for Muslims*), in which the tweeter makes a distinction between religion and extremism. In sum, this cluster contains negative positioning against terrorism and religious extremism.

## 8. Discussion and conclusion

Our paper has focused on a big dataset of a Twitter discussion, using mixed methods. We have aimed at analyzing the topics discussed within one global Twitter discussion related to a news event. The methodology design was executed in a sequential manner: first, we conducted clustering with language technology methods; then, we conducted a quantitative analysis of modality in each cluster on the basis of the most typical tweets; and we finally conducted a digital discourse analysis to identify the stancetaking and positioning of the users participating in this discussion.

According to our first hypothesis, that the #jesuisCharlie Twitter discussion contains multiple topics similar to the discourses that circulated cross-media during the disruptive and hybrid news event that evolved over time ([Chadwick, 2013](#); [Sumiala et al., 2016](#)). We found that the topics could be clustered as news issues for other media (Cluster A), politicians' reactions and actions as well as marches for peace (Cluster B), critical views and defense of freedom of speech (Cluster C), support for freedom of speech and similar events (Cluster D), and Muslims and extremism (Cluster E).

According to our second hypothesis, we found stances that were affective, diversified, and polarized, but in different ways as expected. In certain respects, the topics overlapped in clusters, and the boundaries between clusters were fuzzy. Cluster A contained some tweets whose function was to share information; in this respect, they were the least modalized and affective compared to the others. The analysis of modality showed that Cluster C was different from the others as its tweets were more modalized and thus appeared to be more subjectively positioned than those in the rest of the clusters. Several clusters contained diversified and polarized opinions to some extent, but those with expressions of solidarity (Cluster B) and comments on other people's views (Cluster D) were most diversified and polarized. In other words, counter-discourse was present in these clusters.

The analysis proved to be challenging because of the polarized tendencies described above. Linguistically, the clusters were different, even though they contained some tweets in which opinions seemed to be similar, at least at the surface level. We concluded that critical views of the satire, defense of freedom of speech, and insults to supporters of #jesuisCharlie, as the key topics of Cluster C, inspired a more interactive, argumentatively oriented, and personally involved stance than other topics did.

The *Charlie Hebdo* attacks were a mediatized global news event that garnered reaction all over the world. The anatomy of #jesuisCharlie presented in this article shows varied stances and positionings that users take in the discussion, revealing that the globally networked audience was not unified or like-minded (cf. [Marwick and Boyd, 2011](#)). Participants shared views and understanding of the news event, and some even shared values. In this respect, this Twitter discussion was a self-reflective or "coming together" discourse after a traumatic event, but it was also a description of news taking place and an expression of users' stances regarding freedom of speech, identification and group belonging, and condemnation of terrorism. However, there were tweets that were polemical and criticizing, at places violent, ironical, insulting and bashing counter-discourses

against almost all these sub-discourses (cf. Mercier, 2016). Thus, we describe this #jesuisCharlie massive and global discussion as highly contradictory and controversial.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.pragma.2018.03.007>.

## References

- Badouard, Romain, 2016. « Je ne suis pas Charlie ». Pluralité des prises de parole sur le web et les réseaux sociaux. In: Lefebure, P., Sécaïl, C. (Eds.), *Le défi Charlie. Les médias à l'épreuve des attentats*. Lemieux Editeur, Paris.
- Baroni, Marco, Bernardini, Silvia, Ferraresi, Adriano, Zanchetta, Eros, 2009. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Lang. Resour. Eval.* 43 (3), 209–226.
- Biber, Douglas, Finegan, Edward, 1989. Styles of stance in English: lexical and grammatical marking of evidentiality and affect. *Text* 9 (1), 93–121.
- Bou-Franch, Patricia, Blitvich, Pilar Garcés-Conejos, 2014. Conflict management in massive polylogues: a case study from YouTube. *J. Pragmat.* 73, 19–36.
- Bouko, Catherine, Calabrese, Laura, De Clercq, Orphée, 2017. Cartoons as interdiscourse: a quali-quantitative analysis of social representations based on collective imagination in cartoons produced after the Charlie Hebdo attack. *Discourse Context Media* 15, 24–33.
- Boumédiène, Malik, Catto, Marie-Xavière, Champeil-Desplats, Véronique, Fercot, Céline, Grosbon, Sophie, Gründler, Tatiana, Hennette-Vauchez, Stéphanie, 2017. « Etre Charlie », *La Revue des droits de l'homme*, 7 | 2015. <http://revdh.revues.org/1310> (Accessed 13 February 2017).
- Bruns, Axel, Burgess, Jean E., 2011. The use of Twitter hashtags in the formation of ad hoc publics. In: *Proceeding of the 6th European Consortium for Political Research General Conference (ECPR 2011)*. University of Iceland, Reykjavik, pp. 1–9.
- Bruns, Axel, Highfield, Tim, Burgess, Jean, 2013. The Arab spring and social media audiences: English and Arabic twitter users and their networks. *Am. Behav. Sci.* 57 (7), 87–898.
- Bruns, Axel, Liang, Yuxian E., 2012. Tools and methods for capturing Twitter data during natural disasters. *First Monday* 17 (4). <http://journals.uic.edu/ojs/index.php/fm/article/view/3937/3193>.
- Certeau, Michel de, 1990. *L'invention du Quotidien*. 1. Arts de Faire. Gallimard, Paris.
- Chadwick, Andrew, 2013. *The Hybrid Media System: Politics and Power*. Oxford University Press, Oxford.
- Chiluba, Innocent, Ifukor, Presley, 2015. 'War against our children': stance and evaluation in #BringBackOurGirls campaign discourse on twitter and facebook. *Discourse Soc.* 26 (3), 267–296.
- Creswell, John W., 2013. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage publications, Los Angeles.
- Davies, Bronwyn, Harré, Rom, 1990. Positioning: the discursive production of selves. *J. Theor. Soc. Behav.* 20 (1), 43–63.
- Derczynski, Leon, Ritter, Alan, Clark, Sam, Bontcheva, Kalina, 2013. Twitter part-of-speech tagging for all: overcoming sparse and noisy Data. *Proc. Int. Conf. Recent Adv. Nat. Lang. Process.* 198–206.
- van Dijk, Teun A., 1980. *Macrostructures: an Interdisciplinary Study of Global Structures in Discourse, Interaction, and Cognition*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- DuBois, John, 2007. The Stance Triangle. In: Englebretson, R. (Ed.), *Stancetaking in Discourse. Subjectivity, Evaluation and Interaction*. John Benjamins, Amsterdam/Philadelphia, pp. 139–182.
- Eriksson, Moa, 2016. Managing collective trauma on social media: the role of Twitter after the 2011 Norway attacks. *Media. Cult. Soc.* 38 (3), 365–380.
- Ferenčík, Milan, 2017. I'm not Charlie: (Im)politeness evaluations of the Charlie Hebdo attack in an internet discussion forum. *J. Pragmat.* 111, 54–71.
- Fintel, von Kai, 2006. Modality and language. In: Borchert, D.M. (Ed.), *Encyclopedia of Philosophy*, second ed. MacMillan, Detroit, pp. 20–27.
- Firth, John Rupert, 1957. A synopsis of linguistic theory 1930–1955. *Studies in linguistic analysis*. Oxford: philological society. Reprinted. In: Palmer, F.R. (Ed.), *Selected Papers of J.R. Firth 1952–1959*. Longman, London.
- Giaxoglou, Korina, 2017. JeSuisCharlie? Hashtags as narrative resources in contexts of ecstatic sharing. *Discourse, Context and Media*, 13–20.
- Gries, Stefan Th., 2012. Behavioral Profiles: a fine-grained and quantitative approach in corpus-based lexical semantics. In: Jarema, G., Libben, G., Westbury, C. (Eds.), *Methodological and Analytic Frontiers in Lexical Research*. John Benjamins, Amsterdam & Philadelphia, pp. 57–80.
- Halliday, Michael A.K., 1985. *An Introduction to Functional Grammar*. Arnold, London.
- Halliday, Michael A.K., Matthiessen, C.M.I.M., 2014. *Halliday's Introduction to Functional Grammar*, fourth ed. Routledge, London. revised by C.M.I. Matthiessen.
- Harris, Zellig, 1968. *Mathematical Structure of Language*. Wiley, New York.
- Helasvuoto, Marja-Liisa, Johansson, Marjut, Tanskanen, Sanna-Kaisa (Eds.), 2014. *Kieli verkossa. Näkökulmia digitaaliseen vuorovaikutukseen*. [Language in the Net. Perspectives to Digital Interaction]. SKS, Helsinki.
- Hornik, Kurt, Feinerer, Ingo, Kober, Martin, Buchta, Christia, 2012. Spherical k-means clustering. *J. Stat. Software* 50 (10), 1–22.
- Hunston, Susan, Thompson, Geoff (Eds.), 2003. *Evaluation in Text. Authorial Stance and the Construction of Discourse*. Oxford University Press, Oxford.
- Jaffe, Alexandra, 2009. Introduction: the sociolinguistics of stance. In: Jaffe, A. (Ed.), *Stance: Sociolinguistic Perspectives*. Oxford University Press, Oxford.
- Jones, Karen Spärck, 1972. A statistical interpretation of term specificity and its application in retrieval. *J. Doc.* 28 (1), 11–21.
- Johansson, Marjut, Kleinke, Sonja, Lehti, Lotta (Eds.), 2017. Special Issue: Digital Agora. *Discourse, Context and Media*, 2017, p. 19.
- Jurafsky, Dan, Martin, James H., 2017. *Speech and Language Processing*, third ed. Prentice Hall PTR, NJ, USA.
- Kaufman, Leonard, Rousseeuw, Peter J., 1990. *Finding Groups in Data: an Introduction to Cluster Analysis*. John Wiley, New York.
- Kaufmann, Stefan, Condoravdi, Cleo, Harizanov, Valentina, 2006. Formal Approaches to Modality. In: Frawley, W., Eschenroeder, E., Mills, S., Nguyen, T. (Eds.), *The Expression of Modality*. Mouton de Gruyter, Berlin, New York, pp. 71–106.
- Kiefer, Ferenc, 1987. On defining modality. *Folia Ling.* 21, 67–94.
- Lambiase, Jacqueline J., 2010. Hanging by a thread: topic development and death in an online discussion of breaking news. *language@internet* 7, 9.
- Manning, Christopher, Raghavan, Prabhakar, Schütze, Hinrich, 2008. *Information Retrieval*. Cambridge University Press, Cambridge.
- Martin, James R., White, Peter R.R., 2005. *The Language of Evaluation. Appraisal in English*. Palgrave Macmillan, Basingstoke.
- Marwick, Alice E., Boyd, Danah, 2011. I tweet honestly, I tweet passionately: twitter users, context collapse, and the imagined audience. *N. Media Soc.* 1, 114–133.
- Mercier, Arnaud, 2016. Twitter as a counterpublic sphere. In: Brachotte, G., Frame, A. (Eds.), *Citizen Participation and Political Communication in a Digital World*. Routledge, New York, pp. 139–152.
- Mikolov, Tomas, Chen, Kai, Corrado, Greg, Dean, Jeffrey, 2013. Efficient estimation of word representations in vector space. *Proc. Workshop iCLR*. Retrieved from arXiv preprint arXiv:1301.3781.
- Morozov, Evgeny, 2011. *The Net Delusion: the Dark Side of Internet Freedom*. Public Affairs, New York.
- Nugier, Armelle, Guimond, Serge, 2016. « je suis Charlie » new findings on the social and political psychology of terrorism. *Int. Rev. Soc. Psychol.* 29 (1), 45–49. <https://doi.org/10.5334/irsp.60>.
- Page, Ruth, 2012. The linguistics of self-branding and micro-celebrity in Twitter: the role of hashtags. *Discourse Commun.* 6 (2), 181–201.
- Papacharissi, Zizi, 2015. *Affective Publics. Sentiment, Technology and Politics*. Oxford University Press, Oxford.
- Puschmann, Cornelius, 2015. The form and function of quoting in digital media. *Discourse Context Media* 7, 28–36.

- R Core Team, 2015. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>.
- Scott, Kate, 2015. The pragmatics of hashtags: inference and conversational style on Twitter. *J. Pragmat.* 81, 8–20.
- Smyrniotis, Nikos, Ratinaud, Pierre, 2017. The Charlie Hebdo attacks on twitter: a comparative analysis of a political controversy in English and French. *Soc. Media Soc.* 3 (1), 205630511769364.
- Sumiala, Johanna, Tikka, Minttu, Huhtamäki, Jukka, Valaskivi, Katja, 2016. JeSuisCharlie: towards a multi-method study of hybrid media events. *Media Commun.* 4 (4), 97–108. <https://doi.org/10.17645/mac.v4i4.593>.
- Thompson, Geoff, Hunston, Susan, 2003. Evaluation: an introduction. In: Hunston, S., Thompson, G. (Eds.), *Evaluation in Text. Authorial Stance and the Construction of Discourse*. Oxford University Press, Oxford, pp. 1–27.
- Weizman, Elda, 2008. Positioning in Media Dialogue. John Benjamins, Amsterdam/Philadelphia.
- Whitehead, Tom, 2015. Paris Charlie Hebdo attack: je suis Charlie hashtag one of most popular in twitter history. *Telegraph*, 15.1.2015. <http://www.telegraph.co.uk/news/worldnews/europe/france/11336879/Paris-Charlie-Hebdo-attack-Je-Suis-Charlie-hashtag-one-of-most-popular-in-Twitter-history.html>.
- Yang, Guobin, 2016. Narrative agency in hashtag activism: the case of #BlackLivesMatter. *Media Commun.* 4 (4), 13–17.
- Zappavigna, Michele, 2011. Ambient affiliation: a linguistic perspective on Twitter. *N. Media Soc.* 13 (5), 788–806.
- Zappavigna, Michele, 2014. Enacting identity in microblogging through ambient affiliation. *Discourse Commun.* 2, 209–228.

**Marjut Johansson** is a Professor at the Department of French Studies, at the School of Languages and Translation Studies, University of Turku. Her recent research interest covers a variety of topics on digital interaction and discourse and it includes work that examines online news discourse, social media, digital discussions and videos. At the moment, she is interested in multimodal, digital humanities, and mixed method approaches. She has co-edited several volumes pertaining to these topics: *Language in the Net – Perspectives to Digital Interaction* (Helasvuo et al., 2014), *Digital Agora* (Johansson et al., 2017)

**Aki-Juhani Kyröläinen** is currently a post-doctoral researcher within the Words in the World project. His main research interests center on distributional models of language with a focus on morphosyntactic structures. Additionally, his research combines multiple methodologies ranging from corpus analysis to psycholinguistic experimentation with an emphasis on eye-tracking.

**Attila Krizsán** is a senior lecturer at the English Department of the University of Turku, Finland. His research interests are in functional linguistics and (critical) discourse analysis with a special emphasis on the relationship of language and politics and language and identity. He has published on identification patterns in European political discourse, representation, discourse strategies and political power in Hungary, perceptions of Europeanism and nationalism and networking and multilingualism among EU civil servants and lobbyists and Europeanization of public education systems. He has been working with corpus data since 2003.

**Lotta Lehti**, PhD, is a post-doctoral researcher at the School of Languages and Translation Studies, University of Turku, Finland. Her research interests lie in the fields of discourse analysis and rhetoric and she specializes in digital communication. She has studied, among other themes, ethos in politicians' blogs, flaming in online discussions and argumentation in news sites comment sections.

**Veronika Laippala** is Associate professor of Digital linguistics in the School of languages and translation studies at the University of Turku, Finland. Her research focuses on corpus linguistics and computational linguistics. In particular, she has worked on the development of web-crawled corpora and corpora of computer-mediated communication in various languages and on enhancing computational methods for text linguistics and discourse analysis. Her most recent projects include "Structuring language use across multilingual web corpora" aiming at automatically detecting registers from web corpora and "Whose welfare state?" applying text mining methods to study how welfare state is discussed in social media.

**Filip Ginter** gained MSc (2001) and PhD (2007) in computer science and currently holds the position of an assistant professor in language technology at the University of Turku. His research interests are centered around machine learning applied to large textual corpora.