CORPUS LINGUISTIC STUDY OF TWEETS USING *#CHARLIEHEBDO* HASHTAGS

Intan Siti Nugraha
intan19004@mail.unpad.ac.id
Universitas Padjadjaran
Eva Tuckyta Sari Sujatna
eva.tuckyta@unpad.ac.id
Universitas Padjadjaran
Sutiono Mahdi
sutiono.mahdi@unpad.ac.id
Universitas Padjadjaran

ABSTRACT

Hashtags of *#CharlieHebdo* became trending in Twitter when a knife attack incident happened in front of former Charlie Hebdo magazine's office on 25 September 2020. If *#JeSuisCharlie* was used to show empathy and support for the victim and freedom of speech value, *#CharlieHebdo* still remains question on what topics around the Twitter discussion using the hashtag. Thus, using corpus linguistic analysis method, which are keyword and concordance analysis, this study aims to investigate the significant topic of corpus containing *#CharlieHebdo* hashtag. The tweet corpus which contains 8.604 tweets and retweets and words in total are 177.352 tokens (words) was constructed from the tweets scrapped by the researcher using Python and Twitter API. The result of analysis shows that there are at least 13 categories of keywords which indicate significant topics of the tweet corpus. They are place, attacker, act, weapon, religion/belief, motive, victims, figures, emotion evoked, law enforcement and other topics.

 Keywords: Corpus Linguistics, Keyword Analysis, Concordance analysis, Topics, #CharlieHebdo

INTRODUCTION

Electronic conversation in social networking site (SNS), such as Twitter, has an advance in a way that face-to-face conversation, currently, does not have. They engage in new way of everyday life interaction and discussion with others. For example, electronic conversation in Twitter, through hashtag, facilitates users to search what other people (users) are saying online and to form communities of share value. Hashtag is exploited by user to form interpersonal search function and to form affiliations (De Cock & Pedraza, 2018; Zappavigna, 2012). The affiliations created are likely indirectly through hashtags used to aggregate tweets posted by multiple users who use the same hashtags. This aggregation can create a polyphonic backchannel such as a commentary on a particular event (Reinhardt et al., 2009) both through original posts or comment of news articles attached which gives opportunity for mass participation in the creation, circulation and contestation of discourses (Baker & McEnery, 2015).

The example of the use of hashtags in Twitter showing mass participation in Twitter discourse towards an event is hashtag *#JeSuisCharlie*. Hashtag of *#JeSuisCharlie* becomes a trending and the most popular hashtag event movement in real actions around the world for several period of time especially in the beginning of 2015 when a shooting attack happened in Charlie Hebdo's editorial office on 7th January 2015. In that shooting attack incident, two French muslim brothers killed 12 cartoonists and injured 11 others. They claimed that what they did was the revenge against the magazine which published controversial publication of Prophet Muhammad cartoons in 2011 and 2012 depicted Prophet in nude cartoons. Within hours following the attack, *#JeSuisCharlie* became a trending on Twitter. People used the hashtag to show solidarity and support for the victims and to support freedom of speech value (Giglietto & Lee, 2017; Mondon & Winter, 2017).

In 2020, several years later the attack against Charlie Hebdo's cartoons of Prophet Muhammad, the magazine announce the plan of republish cartoon of Prophet Muhammad on 1 September 2020 ahead of a trial suspected perpetrators of the shooting attack in 2015. In Instagram, two accounts identified as Charlie Hebdo's staffs posted the cartoons. Triggered by similar motive of hatred, on 25 September 2020 a 25-year-old man form Pakistan stabbed and injured two people outside the former of Charlie Hebdo office--near the site of the former Charlie Hebdo office — the scene of a 2015 terrorist attack targeting the satirical newspaper (NYT, 2020). Following the spread of the news around the world, *#CharlieHebdo* became popular on Twitter as the response of the attack on 25 September 2020. The hashtag *#JeSuisCharlie* was not as popular as in 2015 in Twitter discussing an attack targeting Charlie Hebdo magazine which published another cartoon of Prophet Muhammad by its staffs. They used *#CharlieHebdo* instead to mark the topic on discussion of the incident. However, it still remains the question on what the topics around discussion in Twitter *#CharlieHebdo*. If people in Twitter used *#JeSuisCharlie* to show empathy and support for the victim and freedom of speech value, the use of *#CharlieHebdo* still remains question about what topics or shared values are in the Twitter posts using it. For these regards, it is significant to investigate what topics are in the posts using *#CharlieHebdo* hashtag.

Corpus Linguistics

A corpus is a collection of texts that has been compiled for a particular reason based on a set of design criteria, one of which is that the corpus aims to be representative. Biber and Rippen (2015) remark that corpus linguistics is a research which facilitates empirical investigations of language in use and it makes the findings have much greater generalizability and validity. Besides as an empirical research, the characteristics of corpus linguistic analysis are also associated to the use of computerized corpus and tools in analysis and

quantitative and qualitative analytical techniques (Yuliawati et al., 2019; Biber and Rippen, 2015; Partington et al., 2013).

The characteristic of corpus linguistics which is associated to computerized corpus, therefore makes electronically-encoded text such as Twitter posts to be extremely attractive data source. In addition, applying corpus linguistics in analyzing natural language use in Twitter which encompasses large-scale data can contribute to boost empirical credence and to ensure objectivity and full coverage. The statistical significance in the process of analyzing also can increase the level of generality of the research findings and conclusions (Gabrielatos & Baker, 2008) and lend credibility and validity to the analysis. Computational processes of corpus linguistics become the main advantage. Collins (2019) argues that computer can perform better counting and sorting large data in a more accurate, consistent way and much more quickly. It also warrants consistency and minimizes impact of human error and subjective bias of the researcher in the process of analysis.

Existing studies of hashtags have either scrutinized data samples elicited on the basis of searches for tags related to particular events, such as Donald Trump's (Ross & Chaldwell, 2020) and Barack Obama's presidential election (Zappavigna, 2011), the Sydney Seige case (Wendland et al., 2018), an aired documentary television series *Benefits Street* (Baker & McEnery, 2015), kidnapping girls in Nigeria by Boko Haram (Chiluwa & Ifukor, 2015), Schapelle Corby's release day (Zappavigna, 2016). Some studies specifically employ corpus linguistic approach, such as keyword analysis to investigate significant topics and discourses around them (Baker & McEnery, 2015) while other which use focus in news discourse (Al Fajri, 2019; Gabrielatos & Baker, 2008) to make beginning stage of further discourse analysis.

In their study, Baker and McEnery (2015) studied corpus of tweets about to the televised debate of a documentary television series, Benefits Street, broadcasted on 16th February 2014. The tweets collected as the corpus were

posted within a week's worth of tweets, that was from 16th to 23:59 PM on 22nd February 2014. Using corpus linguistics analysis, that is keyword analysis, they investigated the topics around the discussion in the corpus and subsequently carried out more detailed concordance analysis in order to identify discourses. They conducted topic categorization from the keywords by hand using concordance analysis. This keywords analysis reveals what twitter users thought of the debate generally. The result found three main discourses in the corpus along with associated discourse communities which are the idle poor, the poor as victims and the rich get richer with the latter two reinforcing one another (Baker & McEnenry, 2015).

In similar vein with Baker and McEnery (2015), Al Fajri (2019) and Gabrielatos and Baker (2008) investigated keyword classification thematically to find out topics which lead to further discourse analysis. Their results of keyword classification or categorization in their studies were considered to be prominent starting point to follow up its role in a socio-cultural context study in applying corpus-based linguistic study. For this regards, Al Fajri (2019) and Gabrielatos and Baker (2008) provided the methodological framework of corpus-based analysis in which corpus-based analysis such as keyword, concordance and collocation analysis were employed to reveal the frequent topics or issues discussed in news articles discourse. Hence, a question arises as to whether keyword analysis subsequently followed by concordance analysis can elicit topics on Twitter discourse as news articles discourse.

Keywords are words which have a special status because they express important evaluative social meanings and they play a special role in a text or text-type derived from specific statistical process (Bondi & Scott, 2010; Stubbs, 2010). It is called keywords if the word occurs significantly in a text compared against reference corpus (Baker & McEnery, 2015). Technically, it compares word frequency of the corpus with word frequency of reference corpus through statistical probability as computed by an appropriate procedure (log-likelihood

score or the chi-squared test) which is smaller or equal to a p value specified by the researcher (Baker, 2004). Keyword types usually found are proper nouns, keywords commonly recognized as a key, and indicators of the 'aboutness' of a particular text or corpus. Thus, keyword analysis is analysis of significant and frequent words and 'aboutness' of a corpus (Baker, 2006) and when "two corpora are compared together it reveal the most significant lexical differences between them, in terms of 'aboutness' and style (p.347)" (Baker, 2004). Thus

By analyzing corpus of posts to Twitter, this study aims to investigate discourse around incident of knife attack in Paris on 25 September 2020 as well as significant topics in the tweets using hashtag of *#CharlieHebdo*. Corpus linguistic analysis method is fruitful specifically keyword analysis to reveal the 'aboutness' of the corpus. Keywords obtained will indicate the significant topics which are aimed in this research purpose.

METHOD

In accordance with the purpose of the research to find significant topics in the tweet corpus using hashtag of *#CharlieHebdo,* corpus linguistics is employed in term of both quantitative and qualitative analysis methods. It is in line with the characteristic of corpus linguistics which involves both methods. Quantitative analysis helped to handle large-scale of natural language data and qualitative analysis provided a more contextual analysis in this study. In the following subsections, it is elaborated how the corpus as the source of data of this research was constructed and the analysis techniques were used.

Corpus Building

The data were collected through several techniques. First, to get the corpus of tweet intended, the researcher made a program of tweet scraping using Python and Twitter Application Programing Interface (API) to scrape all tweets containing the string *'#CharlieHebdo'*.

All of the scraping procedures using the program made were followed to get the tweets and re-tweets which meet the criteria. The corpus collected containing tweets and retweets were posted within a week, 25 September-1st October 2020, after the incident of Paris knife attack on 25 September 2020. All of the tweets scrapped were limited to the tweets and retweets in English language.

The corpus of tweets collected after 'cleaning up' process contains of 8.604 tweets and retweets and words in total are 177.352 tokens (words). To be available in the software analysis, that is AntConc (Anthony, 2019), the corpus collected from python which is in form of *CSV* format was imported into *.txt* format.

Analytical Framework

For the analysis, keyword and concordance analyses were employed to meet the purpose of the research. Corpus analysis tool AntConc 3.5.8 (Anthony, 2019) was used to derive keyword list and to conduct concordance analysis to the more close-text analysis. Keyword analysis is analysis of 'aboutness' (Baker, 2006) to find keywords in the corpus. A keyword is a word which occurs significantly and frequently in a corpus compared against reference corpus (Anthony & Baker, 2015; Baker & McEnery, 2015; Bondi & Scott, 2010). In this study, to get keyword list of a corpus, the researcher compared a word frequency list of the tweet corpus with WordLex Twitter word frequency list (Gimenes & New, 2015) as the reference. Twitter word frequency list collected by WordLex project (Gimenes & New, 2015) was used as the reference corpus because it was derived from tweet corpus collected in 2012 which contains total words of 30.9 million words.

In conducting keyword analysis in AntConc, the hashtags to take into account for the analysis have been decided to focus only on *#CharlieHebdo* while other hashtags on the corpus will not be taken into account to be known as hashtags, so that it is not necessarily to consider # mark in the analysis.

Meanwhile, the function of @ to mention other user is considered important for keyword analysis to know whether there are figures (users) who become the topic in this corpus. Thus the punctuation of @ is appended to list of user-defined token classes. In addition, the researcher uses loglikelihood with the *p* < 0.001 value regarding to the characteristic of the tweet corpus which is homogenous with reference corpus which is according to Paquot and Bestgen (2009) log-likelihood is appropriate for this kind of corpus.

For more qualitative analysis, 200 top keywords were analyzed using concordance analysis. The context of each keyword were investigated through concordance line in AntConc, This concordance analysis allowed the researcher to conduct a close reading analysis to avoid over and under interpretation of keyword in categorizing topics. Although a tweet which only contains no longer than 140 characters eases the process of concordance analysis, in some cases the researcher sometimes needed to conduct full tweets to avoid the tweets which were not fully depicted in the concordance line display.

FINDINGS AND DISCUSSION

Corpus of tweets using #*CharlieHebdo* was analyzed using keyword and concordance analysis to reveal the significant topics or discourse topics. Keywords derived from AntConc (Anthony, 2019) were categorized thematically and intuitively using quick concordance analysis by hand to determine which group they best belong to. Need to be informed that all the keywords categorized are the content words which are more helpful in indicating topics of the corpus. Stopword list in the process of AntConc keyword analysis helps is generated to produce keywords of content words only. As Bondi & Scott (2010) argue that the relationship of statistical keywords and aboutness is that unusual frequent lexical words differentiates targeting texts or corpus from the

other which consequently indicate the prevailing topic of the targeting texts or corpus. The keywords derived are shown in Table 1 below.

Table-1. Top 200 Keywords

Category	Keywords
Place	paris (4939), france (2153), where (1436), former (2602), near (2110), office (2211), offices (1071), outside (1441), far (639), country (438), Europe (123), arrondissement (101), building (211), area (161)
Attacker	Pakistan (2703), pakistani (1441), suspect (1246), terrorist (264), origin (706), islamists (639), pak (192), culprits (608), muslim (279), terrorists (246), son (246), muslims (142), suspects (211), attackers (171), man (1018), radical, suspect, attacker (145), suspected (177), extremists (161), son (246), born (120),
Act	Attack (3840), terror (449), injured (1083), stabbing (281), attacked (1125), terrorism (768), ongoing (630), stabbed (449), stabbings (132), killed (644), attacks (309), violence (669), massacre (130), threats (125), let (138), knifeattack (120), mindless (610), intimidated (241), upholding

	(241), wounded (274), stepped (186), behead (112), parisattack (162) , hurt (145), actions (108)
Weapon	Knife (1638), machete (686), cleaver (284), gunned (185), meat (239)
Charlie Hebdo Magazine	Cartoons (1953), hebdo (593), charlie (536), cartoonists (612), satirical (417), magazine (483)
Victim	Injured (1083), seriously (253), two (940), four (888), five (293), wounded (274), hand (611), employees (258) victims (123)
Figure	Imrankhan (1369), @imrankanpti (1301), @chrismoored24(399), @emmanuelmacron (367), president (113), minister (150), pm (1378), @mperelman (135), @france_24 (466), prime (121)
Attacker's Motive	Cartoons (1953), prophet (358), muhammad (219), cartoon (180), mohammad (116), drawing (237), blasphemy (101), insult (122)
Emotion	Condemns (1314), condemn (316), islamophobic (1296), islamophobia (263), hatred (122), proud (231), intimidated (241),
Law enforcement	Police (956), arrested (512), detained (135), bastille (167)

Time	Friday (256), timing (240), January (134) years (498)
Religion/belief	Christian (116), islamism (137), islamic (405), freedom (457), value (328), Islam (147),rights (251)
Others	Unga (1377), trial (910), alqaeda (101)

Table 1 shows lists of keywords after being categorized, including names of people which are significant for this study. In determining categories of those keywords, the researcher examined each keyword's context to confirm its theme based on the context of the Paris attack incident on 25 September 2020. For example, *Pakistan* in the keyword list is categorized into "attacker" by the researcher although semantically it is categorized into nation/place. However, after examining the keyword in context in concordance line, the word *Pakistan* refers to the origin of the attacker so that it is classified into the category of "attacker" since the reference is indicating to the same theme of attacker's identity as a 25 year-old immigrant from Pakistan.

From the keyword list, it is seen that keywords categorized into "attacker" and 'the act' frequently occur than other categories. Both categories of 'attacker' and 'the act' are indicated by twenty three keywords which appeared to be the top 200 statistically significant indicating the topic of the corpus. Keywords related to the attacker which mostly refers to his identity have high keyness rank and occurrence such as *Pakistan* (2703) or *Pakistani (1441)* or *pak (192), islamists (639), islamist (401), muslim (279), muslims (142)*. A quick concordance analysis of these words indicates that they reveal his origin as the immigrant from Pakistan and his religion. Mostly, tweets containing those words attach link of news articles exposing the attacker identity and comments on those articles. The examples of tweets are as follows:

(RT) Two people arrested (one just 18 years old) after a stabbing attack close to the former offices of #CharlieHebdo.

—† One of Paris stabbing attackers was born in Pakistan.

Not only attaching the news articles exposing attacker's identity, Twitter users also discuss and put evaluation words consecutively with the word *Pakistan* which generally refers to whole, not as the individual, for example Pakistan as the nation or evaluation for all *Pakistani, islamists* and *muslims*. As depicted to the tweet below, the discussion in of twitter users using those hashtags, in some cases and many of them, expands to more stereotyping the whole Pakistani and Muslim.

"Of course its terrorism since he is a **Muslim**!   ˜'

#CharlieHebdo"

Kundnani (2017) remarks that after the incident of 9/11 in 2001, there has been redefining the concept of extremist and terrorism as the emergence of the so-called 'Global War on Terror' which narrowly referred to specific nation and Islam. Once the act of violence conducted by a people adherent to Islam or Muslim, it will be defined as the act of terror. The other keywords linked to the attacker are words which refer to the people who act violence or criminals such as *terrorist (827), terrorists (264), suspect (1246), culprits (608), attackers (171), attacker (145), extremists (161).* The example of tweet linking the act of violence to Islam is as follows:

"(RT) In the space of 3. days: Meat cleaver attacks in Paris. A MEAT CLEAVER! **MUSLIM** attacker. Police officer shot dead in London. MUSLIM attacker. But we mustn't comment on the constant stream of **ISLAMIC** violence, lest we OFFEND someone. #CharlieHebdo #ParisAttacks #Croydon"

Besides the keywords indicating attacker, keywords categorized into the act are also dominant in the corpus. Twitter users tend to use various words to refer to what happened in the incident such as *attack (3889), terror (2182), injured (1942), stabbing (1689), terrorism (788), violence (672), killed (645), mindless (610), wounded (274), threats (184), threat (166), massacre (130),* and other words which the actor is the attacker. The act category of keywords indicates that topic of discussion are he 'how it happened'. Similar to the category of 'attacker', Twitter users mostly attach news articles in their tweet to be compliment of their discussing commenting on the incident using hashtag of *#CharlieHebdo*

> *(*RT) Just in | Four people were injured, two seriously, in a knife attack in Paris Friday outside the former offices of French satirical magazine #CharlieHebdo, Prime Minister Jean Castex said, police saying one suspect had been detained after the attack.
> - AFP"

Twitter users also use those words subsequently with the attacker and his identity. This is shows the correlation of his identity as Pakistani and Muslim is strongly attached to the violence or terror acts. Therefore, as a means of analyzing further, the researcher conducted a close investigation on the keyword in context (KWIC) in concordance line of AntConc in order to find out how they are related each other. The result finds many of the violence or terror acts largely discussed and related to Pakistan/Pakistani and Islam/Muslim generally.

> "(RT) Paris **Knife attack** perpetrator near old #CharlieHebdo office is a Pakistani.
> **#Pakistan**, exporting terrorists since 1947
> @FATFNews
> Are you watching?"

The tweet above shows how Twitter users link the violence act of incident to Pakistan as a nation. They frame Pakistan with the terrorism using the fact that the attacker in front of former Charlie Hebdo's office is an immigrant from Pakistan. There are also found similar tweets that have same pattern relating the act of terrorism with either Pakistan or Islam generally. Moreover, many the tweets relate the acts with both Pakistan and Islam as instance in follow:

"(RT) Here, a **Pakistani** involved in **knife attack** in #Paris at old office of #CharlieHebdo

Pakistan fountainhead of **Islamic** terrorism

Subsequently with the 'attacker' and 'act' categories, the keywords indicating 'place' category also appear to be significant. This is in line with the appearance of 'attacker' and 'act' keywords which also mainly as comments of the incident news articles. It obviously makes the words, such as *paris (4939), France (2153), former (2602), office (2211)* and other words indicating the place or site of incident to be significant topic in the Twitter. Here is the example of the tweets:

(RT) Charlie Hebdo knife attack breaking: Paris police say a suspect believed to have wounded four people in stabbing near the **former offices** of satirical newspaper #CharlieHebdo has been arrested.

Initially police thought it was 2 men, they now believe it was only one. - #France   (link) "

The next obvious result of the keyword category is 'attacker's motive' which also significant in the corpus. The words indicated attacker's motive such *Cartoons (1953), prophet (358), muhammad (219), cartoon (180), mohammad*

(116), trial (910), drawing (108), blasphemy (101), insult (122) and other words which after being investigated in the concordance line refer to the reason the attacker committed the action. In addition, keywords categorized as 'religion/belief' also prominent as the indicator of topics. How Twitter users discussed the incident expand to the discussion of specific religion, belief, and value.

> (RT) I fully condemn today's Islamist terrorist attack in #Paris. We stand with #France and will not be intimidated, in particular when it comes to upholding our values and fundamental rights, such as free speech. #CharlieHebdo @EmmanuelMacron"

> Or
> "(RT) I honestly don't get Muslim fundamentalist. If you **HATE** western values and ideals so much, then pack your shit and move to saudi arabia or any other sharia **islamic** country. Don't live in western countries and expect them to change their ways for your religious crap. #CharlieHebdo"

The discussion emotion towards particular religion also frequently tweeted by the users as displayed in the example above. The word *Islamic (405)* mostly refers to the word *attack* (161 times as collocate) which is primed as an 'islamic' ways or an act allowed in Islam. This draws the conclusion that each of the categories of keywords is linked in particular ways.

CONCLUSIONS

The research is aimed to reveal the significant topics or discourse topic of the tweets using *#CharlieHebdo* hashtag. Corpus linguistic analysis methods such as keyword and concordance analysis were employed since they are fruitful specifically keyword analysis to reveal the 'aboutness' of the corpus. Keywords

obtained will indicate the significant topics which are aimed in this research purpose. The result shows that there are at least 13 categories of keywords which indicate significant topics of the tweet corpus containing *#CharlieHebdo* hashtag. The significant topics identified are topic related to the fact of the incident such as the place, weapon and time of the incident, topic of the attacker including his identity, origin and himself as the actor of the incident, topic related to the act committed by the attacker, topic of religion, the victims, Charlie Hebdo magazine, emotion evoked by the incident, attacker's motive and topic related to the law enforcement. Those categories however are connected each other in which for example the Twitter users mostly do not only discuss the attacker as the suspect but also link it with his religion that is Islam. Many of the users also discuss Pakistan and Muslim generally framed by this incident discussion.

REFERENCES

Al Fajri, M. S. (2019). The discursive portrayals of Indonesian Muslims and Islam in the American press: A corpus-assisted discourse analysis. *Indonesian Journal of Applied Linguistics*, *9*(1), 167-176.

Anthony, L. (2019). AntConc (Version 3.5. 8) (Computer Software), Waseda University, Tokyo.

Anthony, L., & Baker, P. (2015). ProtAnt: A tool for analysing the prototypicality of texts. *International Journal of Corpus Linguistics*, *20*(3), 273-292.

Baker, P. (2004). Querying keywords: Questions of difference, frequency, and sense in keywords analysis. *Journal of English Linguistics*, *32*(4), 346-359.

Baker, P. (2006). *Using corpora in discourse analysis*. A&C Black.

Baker, P. & McEnery, T. (2015). Who Benefiits When Discourse Get Democratised? Analyzing a Twitter Corpus around the British Benefits Street Debate.

Bondi, M., & Scott, M. (Eds.). (2010). *Keyness in texts* (Vol. 41). John Benjamins Publishing

Biber, D., & Reppen, R. (Eds.). (2015). *The Cambridge handbook of English corpus linguistics*. Cambridge University Press.

Chiluwa, I., & Ifukor, P. (2015). 'War against our Children': Stance and evaluation in #BringBackOurGirls campaign discourse on Twitter and Facebook. *Discourse & Society*, *26*(3), 267-296.

Collins, L. C. (2019). *Corpus linguistics for online communication: A guide for research*. Routledge.

De Cock, B., & Pedraza, A. P. (2018). From expressing solidarity to mocking on Twitter: Pragmatic functions of hashtags starting with# jesuis across languages. *Language in society*, *47*(2), 197.

Gabrielatos, C., & Baker, P. (2008). Fleeing, sneaking, flooding: A corpus analysis of discursive constructions of refugees and asylum seekers in the UK press, 1996-2005. *Journal of English linguistics*, *36*(1), 5-38.

Giglietto, F., & Lee, Y. (2017). A hashtag worth a thousand words: Discursive strategies around# JeNeSuisPasCharlie after the 2015 Charlie Hebdo shooting. *Social Media+ Society*, *3*(1), 2056305116686992.

Gimenes, M., & New, B. (2015). Worldlex: Twitter and blog word frequencies for 66 languages. *Behavior research methods*, *48*(3), 963-972.

Kundnani, A. (2017). Extremism, Theirs and Ours: Britain's Generational Struggle'. *After Charlie Hebdo: Terror, Racism and Free Speech. London: Zed Books. A 'Muslim'response*, *193*.

Mondon, A., & Winter, A. (2017). Charlie Hebdo, Republican Secularism and Islamophobia

Paquot, M., & Bestgen, Y. (2009). Distinctive words in academic writing: A comparison of three statistical tests for keyword extraction. In *Corpora: Pragmatics and discourse* (pp. 247-269). Brill Rodopi.

Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies (CADS)* (Vol. 55). John Benjamins Publishing.

Reinhardt, W., Ebner, M., Beham, G., & Costa, C. (2009). How people are using Twitter during conferences. *Creativity and Innovation Competencies on the Web. Proceedings of the 5th EduMedia*, 145-156.

Ross, A. S., & Caldwell, D. (2020). "Going negative": An Appraisal analysis of the rhetoric of Donald Trump on Twitter. *Language & Communication*. doi:10.1016/j.langcom.2019.09.003

Wendland, J., Ehnis, C., Clarke, R. J. & Bunker, D. (2018). Sydney siege, December 2014: A visualisation of a semantic social media sentiment analysis. In K. Boersma & B. Tomaszewski (Eds*.), Proceedings of the 15th ISCRAM Conference* (pp. 493-506).

Yuliawati, S., Dienaputra, R. D., Sujatna, E. T. S., Suryadimulya, A. S., & Lukman, F. (2019). Looking into "Awewe" and "lalaki" in the Sundanese Magazine Mangle: Local Wisdom and a Corpus Analysis of the Linguistic Construction of Gender. *International Journal of Advanced Science and Technology*, *28*, 549-559.

Zappavigna, M. (2011). Ambient affiliation: A linguistic perspective on Twitter. *New media & society*, *13*(5), 788-806.

Zappavigna, M. (2012). *The Discourse of Twitter and Social Media (Continuum Discourse Series)*. Continuum International Publishing Group Limited.

Zappavigna, M. (2016). Searchable talk: The linguistic functions of hashtags in tweets about Schapelle Corby. *Global Media Journal. Australian Edition*, (10).

Zubiaga, A. (2018). A longitudinal assessment of the persistence of twitter datasets. *Journal of the Association for Information Science and Technology*, *69*(8), 974-984.