The 5g conspiracy and Covid-19: How language affects the spreading of misinformation on Twitter

Marco Visentin
University of Bologna
Annamaria Tuan
University of Bologna
Giandomenico Di Domenico
University of Porthmouth - Faculty of Business and Law

Cite as:
Visentin Marco, Tuan Annamaria, Di Domenico Giandomenico (2021), The 5g conspiracy and Covid-19: How language affects the spreading of misinformation on Twitter. *Proceedings of the European Marketing Academy*, 50th, (94492)

Paper from the 50th Annual EMAC Conference, Madrid, May 25-28, 2021

The 5g conspiracy and Covid-19: How language affects the spreading of misinformation on Twitter

Abstract

During the COVID-19 pandemic, the spreading of misinformation has reached unprecedented peaks on social media platforms. The analysis of language is of paramount importance in understanding the motivational determinants of social media users in sharing misinformation, as language style conveys meanings and affects the engagement to social media contents. In this study, we analyze textual and non-textual cues from 4923 tweets containing the hashtags #5G and #Huawei published during the first week of May 2020, when conspiracy theories linking the 5G technology to the spreading of the pandemic were circulating online. Through logistic and Poisson regression approaches, we evaluate the effects of the textual and non-textual cues on the retweeting rates. In particular, we find that textual cues play a more important role than the veracity of contents in the spreading of misinformation on Twitter. Moreover, the role of confirmation bias in eliciting misinformation sharing is confirmed.

Keywords: Misinformation, Twitter, linguistic analysis.

Track: *Digital Marketing & Social Media*

1. Introduction

The COVID-19 pandemic has proven to be not only an epidemic but also an "infodemic" given that the spreading of the virus has been accompanied by an explosion of misinformation about the disease (WHO, 2020). During the last months, the 5G conspiracy has spread massively on social media, becoming a trending topic on Twitter and arousing anger among users (Jolley & Paterson, 2020). As a consequence, several cell phone masts have been vandalised in the United Kingdom (BBC, 2020) and the reputation of specific Chinese technology companies, such as Huawei, that have been investing in the development of the 5G network has been tarnished (Di Domenico & Visentin, 2020). Together with social media algorithmic characteristics (Di Domenico et al., 2020), the language that social media users adopt to communicate meanings (Hewett et al., 2016) can be a factor in fostering the spreading of misinformation. On Twitter specifically, the choice of the words used and the hashtags selected enables or restricts the possibility of spreading contents within *echo chambers*, digital environments where misinformation can thrive (Quattrociocchi et al., 2016; Hewett et al., 2016). A growing body of literature investigates Twitter from the consumer point of view (e.g. Kim & Song, 2016) and only recently the influence of the linguistic style of tweets on retweets rates has been more deeply investigated (e.g. Aleti et al., 2019; Labrecque et al. 2020). The analysis of Twitter user-generated textual data can provide unprecedented possibilities to analyse how individuals use language to communicate misinformation on social media and boost contents' visibility.

To this aim, we retrieved a set of 4923 tweets containing the hashtags #5G and #Huawei during the first week of May 2020. We evaluated the impact of textual cues extracted using Linguistic Inquiry and Word Count (LIWC) and non-textual cues on retweet rates.

2. Fake news and misinformation

Fake news is defined as news "intentionally and verifiably false and could mislead readers" (Allcott & Gentzkow, 2017, p. 213). Fake news and other forms of misinformation (e.g. hoaxes, propaganda and conspiracy theories) are nothing new (Tandoc et al., 2018), but they have found in social media a powerful medium to spread broadly and magnify their reach. Indeed, previous research has found that fake news travel faster and farther on social media platforms (Vosoughi et al., 2018).

To date, most of the literature on fake news has tried to understand the drivers underpinning fake news belief and sharing on social media. One of the most important drivers of belief in fake news was found to be confirmation bias (Quattrociocchi et al., 2016), that is *"the human tendency to acquire information adhering to one's system of beliefs"* (Del Vicario et al., 2019, 10:2). Confirmation bias explains the human tendency to select information consistently with prior beliefs, thus reinforcing them (Lewandowsky et al., 2012). Cognitive abilities also play an important role in fostering individuals' belief in fake news (Pennycook et al., 2018). Emotions also can prompt belief in fake news. According to Martel and colleagues (2020), indeed, reliance on emotions increases the susceptibility to fake news.

By mainly focusing on the psychological drivers of belief in and sharing of fake news, existing scholarship on this intriguing topic leaves some gaps in the understanding of the textual characteristics that make fake news achieve virality. As individuals create and extract meanings from texts (Marwick, 2018), the analysis of language is crucial in the comprehension of misinformation sharing behaviours. Moreover, the use of specific words or hashtags can determine the overall reach of the content on social media, stimulating the reverberating system of the "*echoverse*" (Hewitt et al., 2016). For this study, we focus on Twitter for two reasons. Firstly, the spreading of misinformation through this platform is pervasive and it has been chosen as the context of several misinformation studies (e.g. Jung et al., 2018; Vosoughi et al., 2018). Secondly, the text of tweets constitutes a valuable starting point to study how language is built (Berger et al., 2020) and affect others (by the means of retweets). As a consequence, it becomes relevant answering to the following research questions: RQ1: *Which are the main tweets' cues that lead a tweet to be retweeted?* And: *RQ2: Which are the main tweets' cues that increase the virality of tweets?*

3. Textual and non-textual cue stimulating retweets

In order to answer our research questions, we draw upon cueing theory, which suggests that visual and verbal stimuli affect different levels of individual responses (e.g. Labracque et al. 2020; Visentin & Tuan, 2020). In the context of Twitter, cues can affect consumers' sharing behavior in terms of retweets. We focus on textual cues represented by the *linguistic style* of the text, the level of *certainty* of the statements, the *complexity* of the text and the presence of *misinformation*. We complement our framework by considering also non-textual cues such as the *status of the user* and the *use of media*. Figure 1 depicts our theoretical framework.

--- Insert Figure 1 here ---

3.1. Textual cues: linguistic style, certain language, complexity and misinformation

Even though the marketing literature has traditionally focused on the content of social media posts, mainly from a company perspective, in recent years the attention has moved towards the *linguistic style* used in the text suggesting that content words and function words are equally important to communicate meaning, to develop mental imagery, to direct the attention of the reader as well as to affect engagement (e.g. retweets) (e.g. de Vries et al., 2012; Douglas et al., 2017; Aleti et al, 2019; Labracque et al., 2020). Aleti et al. (2019) analyzed, the narrative/analytical, internally/externally focused (*clout)*, and negative/positive emotional styles (*tone*) in tweets by celebrities and their effect on word of mouth. Berger et al. (2020) found that tweets containing a more analytical language—but not authentic language style—were more likely to be shared after the debate. Finally, Humphreys et al. (2020) found that consumers use more abstract (concrete) language in their online search during the first (last) stages of the shopping journey.

Complexity refers to the presence in texts of a high *number of words, propositions*, *long* words and more words related to *cognitive mechanisms* (Tausczik & Pennebaker, 2010). In the case of misinformation, text *complexity* may be reduced because of the cognitive load required to maintain a story that is contrary to experience. Against a scant literature, we expect that the more a text is complex, the less it will be convincing and that, in turn, this text will be shared less. *Certainty* refers to a sense of conviction or confidence that characterizes language. Previous research suggests that certain language increases consumer engagement to brands' social media messages (Pezzutti et al., 2021).

Furthermore, *misinformation* has found a fertile ground on social media, and in particular on Twitter, allowing people to disseminate unreliable information and misleading content (Di Domenico et al., 2020; Del Vicario et al., 2019). However, when this content aligns with an individual's perception of the world, the possibility of being shared increases irrespective of factual truthfulness and objective reality, affecting therefore the virality of the tweet.

3.2. Non textual cues: user status and use of media

In addition to the content and the style of the tweet, we also consider cues that are peripheral to the text, like the *user status* and the *use of media*. First, previous literature suggests that the number of followers, the number of friends and the volume of statuses act as an indicator of source influence for the reader (Xu & Zhang, 2018). These indicators shape the *user status* and represent cues which capture the attention of the user, providing information about the level of authority of the user profile. Second, the inclusion of images or URLs characterizes the levels of vividness and interactivity of the text, which may influence virality as non-textual cues representing the *use of media* (de Vries et al., 2012; Tellis et al. 2019).

4. Empirical analysis

In this study, we focus on the diffusion of misinformation regarding the company Huawei being associated with false rumours regarding insidious connections between the spread of coronavirus and the deployment of the 5G network. To this aim, we scraped from Twitter 4923 tweets containing the hashtags #5g #Huawei on May, 8 2020. The dataset was streamlined by removing verified users and retweets, resulting in 1103 tweets. First, we manually coded tweets to scrutinize whether they were related to *misinformation* or not. We found that 223 out of 1103 tweets were misinformative. Then, we performed the automated text analysis of tweets by using LIWC (Linguistic Inquiry and Word Count; Tausczik & Pennebaker, 2010).

We used the summary variables provided by LIWC to account for *analytical language, authentic language, clout* and *tone,* validated by previous research (Pennebaker et al. 2015; Aleti et al., 2017). *Analytic* captures the narrative/analytical style, characterized by the presence of more exclusive words (e.g. but, while, whereas), more self-and-other references, and less negative emotion denotes an *analytical language*. *Authentic* combines the positive loading of first- and third-person singular pronouns, third-person plural pronouns, and exclusive words (e.g., but, except, and without) with the negative loading of negative emotions and motion verbs (e.g., arrive, drive, and go). Conversely, low scores on this variable relate to more distance from the self and a more deceptive language. *Clout* refers to the degree to which texts contain an internally or externally focused style. *Tone* captures the negative (e.g., hurt, nasty, ugly) and positive (e.g., love, nice, sweet) emotional style.

To account for *text complexity*, we included the following categories of LIWC: word count (*Word Count*), presence of prepositions (*Prepositions*), words with more than six letters (*Long words*) and cognitive mechanisms (*Cognitive Mecs*). To account for *certain language*, we included words communicating possibility (*Tentativeness*) and certainty (*Certainty*).

4.1. Results

Logit Analysis: In order to answer to our first research question, we estimated a Logit model on the retweet of the 1103 tweets. We used the dummy variables *Retweet/No retweet* as the dependent variable, *Analytic, Authentic, Clout, Tone, complex language, certain language* and *misinformation* as independent variables. We also added the dummy variable *Misinformation*, the user's characteristics and the presence of media elements as controls. The results of the Logit analysis on the possibility of a tweet to be retweeted are displayed in Table 1.

--- insert Table 1 here ---

Results indicate a significant effect only of the variables related to the complexity of the language, namely the word count (*Word Count*), words with more than six letters (*Long Words*) and cognitive mechanisms (*Cognitive Mecs*), suggesting that longer and more complex tweets are retweeted more. Meanwhile, *Misinformation* does not report a significant value meaning that the retweet does not depend on the veracity of the tweet but rather on other textual and non-textual cues. Finally, a significant and positive effect of the presence of URLs suggests that when the tweet is complemented by a link to external sources it catches more attention by the reader, eliciting retweeting behaviors. Overall, these results indicate that the presence of complex language and links to external sources in tweets positively affect the probability of a tweet being retweeted.

Poisson analysis: To answer to our second research question, we estimated a Poisson model on the retweet count for the 175 retweets in our dataset. We accounted for the same independent variables of the previous model but we used the *retweet count* as the dependent variable. Results are displayed in Table 2 (differences between the full model and all the sub-models are significant ($P(\chi^2, df)<1e\text{-}03$).

--- insert Table 2 here ---

A different picture is provided by modelling the ability of a tweet to accumulate retweets. We found a positive and significant effect of *Analytic*, *Authentic*, *Clout* and a negative effect of *Tone.* Thus, tweets are more retweeted when they contain a more analytical, authentic and confident language. Whereas negative tweets are less appreciated by the users. The significant effect of the *complexity* variable is maintained and the presence of *Misinformation* is significant. Interestingly, in this case, also the non-textual cues play an important role in determining the virality of tweets. In particular, the *Followers Count* and the volume of tweets (namely, *Statuses Count*) have a positive and significant effect on the *retweet count*. *Friends count* and the presence of images (namely, *Media*) have a negative effect, instead.

5. Conclusion

This study, we analyzed a panel of 4923 tweets and found different effects of the textual and non-textual cues on the retweeting of a tweet and on its ability to accumulate retweets. In particular, the misinformation included in a tweet plays an interesting role in spreading the tweet through the network. More importantly, the relative influence of the cues suggests that Twitter users actually read a tweet but not necessarily they understand or critically evaluate it before deciding to share it on the platform, supporting the role of confirmation bias in affecting individuals' susceptibility to misinformation as the tendency to propagate contents through social media largely depends on other textual cues than the veracity of contents.

We contribute to the nascent stream of literature regarding fake news from a marketing perspective which is still at a nascent stage (e.g. Visentin et al. 2019; Di Domenico & Visentin, 2020; Di Domenico et al. 2020), provinding evidence about the textual cues in affecting retweeting behavior. These cues are not only related to the content of the tweet but also to the style used. Second, we contribute to the stream of literature which aims to analyze the features of tweets affecting virality (e.g. Berger et al., 2020), answering to recent calls about the importance of providing models using language analysis and machine learning techniques (Valsesia et al., 2020).

From a managerial perspective, this study suggests that companies should continuously monitor tweets which are going to become viral in order to avoid the spread of misinformation to avoid negative impact on the company's reputation. Indeed, results suggest that users read tweets before retweeting them but it is not clear whether they have understood them or not. This study suggests that by detecting misinformation, platforms should not only focus on the content (i.e. misinformation) but also on how it is conveyed.

References

Ahmed, W., Vidal-Alaball, J., Downing, J., & Seguí, F. L. (2020). COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. *Journal of Medical Internet Research*, *22*(5), e19458.

Aleti, T., Pallant, J. I., Tuan, A., & van Laer, T. (2019). Tweeting with the stars: Automated text analysis of the effect of celebrity social media communications on consumer word of mouth. *Journal of Interactive Marketing*, 48, 17-32.

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, *31*(2), 211-36.

BBC. (2020, April 23). Coronavirus: "Murder threats" to telecoms engineers over 5G. https://www.bbc. co.uk/news/newsbeat-52395771

Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. Journal of Marketing, 84(1), 1-25.

Brewis H. Evening Standard. 2020 Apr 14. Nightingale hospital phone mast attacked as 5G conspiracy theory rages on URL: *https://www.standard.co.uk/news/uk/nhs-nightingale-phone-mast-arson-attack-5g-conspiracy-a4414351.html*

De Vries, L., Gensler, S., & Leeflang, P. S. (2012). Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing. Journal of interactive marketing, 26(2), 83-91.

Del Vicario, M., Quattrociocchi, W., Scala, A., & Zollo, F. (2019). Polarization and fake news: Early warning of potential misinformation targets. *ACM Transactions on the Web (TWEB)*, *13*(2), 1-22.

Di Domenico, G., Sit, J., Ishizaka, A., & Nunan, D. (2020). Fake news, social media and marketing: a systematic review. *Journal of Business Research*, *Forthcoming.*

Di Domenico, G., & Visentin, M. (2020). Fake news or true lies? Reflections about problematic contents in marketing. *International Journal of Market Research*, 1470785320934719.

Douglas, K. M., Sutton, R. M., & Cichocka, A. (2017). The psychology of conspiracy theories. *Current directions in psychological science*, *26*(6), 538-542.

Hewett, K., Rand, W., Rust, R. T., & Van Heerde, H. J. (2016). Brand buzz in the echoverse. *Journal of Marketing*, 80(3), 1-24.

Humphreys, A., Isaac, M. S., & Wang, R. J. H. (2020). Construal Matching in Online Search: Applying Text Analysis to Illuminate the Consumer Decision Journey. Journal of Marketing Research, 0022243720940693.

Jolley, D., & Paterson, J. L. (2020). Pylons ablaze: Examining the role of 5G COVID‐19 conspiracy beliefs and support for violence. *British journal of social psychology*, *59*(3), 628-640.

Kim, J., & Song, H. (2016). Celebrity's self-disclosure on Twitter and parasocial relationships: A mediating role of social presence. *Computers in Human Behavior*, *62*, 570-577.

Knuutila, A., Herasimenka, A., Au, H., Bright, J., & Howard, P. N. (2020). Covid-Related Misinformation on YouTube.

Labrecque, L. I., Swani, K., & Stephen, A. T. (2020). The impact of pronoun choices on consumer engagement actions: Exploring top global brands' social media communications. *Psychology & Marketing*, 37(6), 796-814.

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological science in the public interest*, *13*(3), 106-131.

Martel, C., Pennycook, G., & Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. *Cognitive research: principles and implications*, *5*(1), 1-20.

Marwick, A. E. (2018). Why do people share fake news? A sociotechnical model of media effects. *Georgetown Law Technology Review*, 2(2), 474–512.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. Personality and social psychology bulletin, 29(5), 665-675.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015.

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of experimental psychology: general*, *147*(12), 1865.

Quattrociocchi, W., Scala, A., & Sunstein, C. R. (2016). Echo chambers on Facebook. *Available at SSRN 2795110*.

Tandoc Jr, E. C., Lim, Z. W., & Ling, R. (2018). Defining “fake news” A typology of scholarly definitions. *Digital journalism*, *6*(2), 137-153.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, *29*(1), 24-54.

Tellis, G. J., MacInnis, D. J., Tirunillai, S., & Zhang, Y. (2019). What drives virality (sharing) of online digital content? The critical role of information, emotion, and brand prominence. Journal of Marketing, 83(4), 1-20.

Valsesia, F., Proserpio, D., & Nunes, J. C. (2020). The Positive Effect of Not Following Others on Social Media. Journal of Marketing Research, 0022243720915467.

Visentin, M., & Tuan, A. (2020). Book belly band as a visual cue: Assessing its impact on consumers’ in-store responses. Journal of Retailing and Consumer Services, 102359.

Visentin, M., Pizzi, G., & Pichierri, M. (2019). Fake news, real problems for brands: The impact of content truthfulness and source credibility on consumers' behavioral intentions toward the advertised brands. *Journal of Interactive Marketing*, *45*, 99-112.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146-1151.

WHO (2020). Munich Security Conference, 15 February 2020, https://www.who.int/dg/speeches/detail/munich-security-conference

Xu, W. W., & Zhang, C. (2018). Sentiment, richness, authority, and relevance model of information sharing during social Crises—the case of# MH370 tweets. Computers in Human Behavior, 89, 199-206.

Figure 1. Conceptual framework

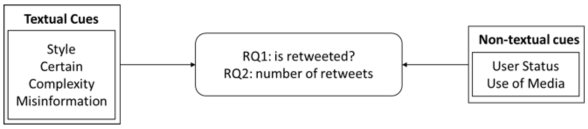

Table 1: Logit Analysis

Linguistic Style		Certain Language		Complexity		Misinformation		Non-textual cues	
Analytic	.140 (.149)	Tentativeness	.024 (.110)	Word Count	.311*** (.094)	Misinformation	.222 (.220)	Followers Count	.125 (.082)
Authentic	.095 (.103)	Certainity	.145 (.108)	Prepositions	.137 (.109)			Friends Count	.202 (.180)
Clout	-.054 (.103)			Long Words	.216** (.095)			Statuses Count	-.071 (.098)
Tone	.117 (.087)			Cognitive Mecs	-.261** (.132)			Urls	.501** (.213)
								Media	.186 (.251)

P(χ^2, df)<1e-03; Observations: 1,103; Log Likelihood: -450.910; AIC: 935.819; Intercept: -2.168***(.180)

*Note: *p<0.1; **p<0.05; ***p<0.01*

Table 2: Poisson Analysis

Linguistic Style		Certain Language		Complexity		Misinformation		Non-textual cues	
Analytic	.109** (.051)	Tentativeness	.002 (.057)	Word Count	.276*** (.036)	Misinformation	.825*** (.081)	Followers Count	.539*** (.039)
Authentic	.105*** (.038)	Certainity	.073 (.045)	Prepositions	.013 (.044)			Friends Count	-.491*** (.048)
Clout	.363*** (.042)			Long Words	.103*** (.039)			Statuses Count	.228*** (.034)
Tone	-.104*** (.040)			Cognitive Mecs	-.102* (.052)			Urls	-.106 (.084)
								Media	-.275*** (.099)

P(χ^2, df)<1e-03; Observations: 175; Log Likelihood: -857.673; AIC: 1,749.346; Intercept: 1.248*** (.076)

*Note: *p<0.1; **p<0.05; ***p<0.01*