Lingue e Linguaggi
Lingue Linguaggi 47 (2022), 191-223
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v47p191
http://siba-ese.unisalento.it, © 2022 Università del Salento

A MIXED-METHOD CORPUS APPROACH TO THE COVID-19 VACCINATION DEBATE

CLAUDIA ROBERTA COMBEI
UNIVERSITÀ DEGLI STUDI DI PAVIA

Abstract – Social media have contributed to the recent proliferation of online discussions on the COVID-19 vaccines. The paper explores the evolution of this debate by analysing an *ad hoc* corpus of tweets (over 5.5 million words) collected from March 15th to April 14th, 2021. We deploy sentiment, emotion, and emoji analysis to uncover the users' affective states, perceptions, and reactions regarding the COVID-19 vaccination. Our results show that vaccine sentiment is influenced by real-time news and by other information that circulates on the Internet, displaying polarizations on both the negative and the positive extremities of the sentiment scale. The emotion analysis indicates that *trust* issues (either *trust* or *mistrust*) regarding the COVID-19 vaccination prevail in our data, amounting to 21.29% of the overall emotional valence of tweets. Furthermore, the qualitative analysis suggests that the infodemic relies primarily on strong negative emotions (e.g., *fear*, *anger*, and *disgust*). Finally, the emoji analysis reveals that, besides iconicity functions, emoji act as boosters of emotions, contributing to the semantic dimension of the Twitter debate on the COVID-19 vaccination.

Keywords: sentiment analysis; emotion analysis; emoji analysis; misinformation; COVID-19 vaccine.

1. Introduction

The scholarly debate has suggested that an insufficient COVID-19 vaccination coverage is problematic as it may retard or hamper the post-pandemic recovery (Lazarus et al. 2021). When vaccines are available, a suboptimal vaccination coverage is generally caused by vaccine hesitancy (Kang et al. 2017). By January 10th, 2022, 5.5 million people died from COVID-19 and over 307 million infection cases were reported[1]; for this reason, the COVID-19 vaccination campaign, in general, and the vaccine hesitancy, in particular, are topics that hold the attention of institutions and organizations from all around the world, and scholars working in various fields of research. For instance, in a recent interdisciplinary work, de Figueiredo and Larson (2021) explore how the propensity to accept a COVID-19 vaccine varies from a geographical and

[1] Up to date information on the COVID-19 cases is available online at this website: https://coronavirus.jhu.edu/map.html (10.1.2022).

a socio-demographical point of view. The results of their survey unveil that the respondents from Lebanon, France, Croatia, and Serbia are less determined to accept a COVID-19 vaccine. In other respects, being male, older, or having a high level of education is associated with a higher likelihood to agree to the COVID-19 vaccination.

Existing studies on COVID-19 vaccine hesitancy mention safety concerns, the rapid pace of vaccine development, the accelerated approval process, and misinformation, as primary reasons of scepticism (Machingaidze and Wiysonge 2021; Wouters et al. 2021). On the same note, Lyu et al. (2021) explore social media in order to understand the public opinion on COVID-19 vaccines. The authors employ a human-guided machine learning approach to investigate the opinions of over 10,000 Twitter users with respect to COVID-19 vaccines. Their system classifies the users into three groups: pro-vaccine, vaccine-hesitant, and anti-vaccine. The results of the study reveal that religious people and socio-economically disadvantaged groups are more likely to display polarized opinions on COVID-19 vaccines – either pro-vaccine or anti-vaccine. Moreover, people living in suburban or rural areas and those who have had the worst personal pandemic experience are more likely to have an anti-vaccine opinion.

At the time this paper was written, few works in the field of linguistics focused on this matter. An important contribution is Breeze's (2021) corpus-assisted discourse analysis of online comments to the Mail Online articles on the development of COVID-19 vaccines. The author explains that the constant demand for health news has led to a huge availability of information from official sources, from the traditional media, and from user-generated online postings. As Breeze (2021) points out, the lattermost are generally viewed as having a tendency to spread misinformation or other harmful information, while, at the same time, the "expert" knowledge is constantly questioned by the general public. Besides the afore-mentioned study, the newly launched Quo VaDis project[2] (coordinated by Elena Semino, at the University of Lancaster) uses corpus-based discourse analysis techniques to explore vaccination concerns and to analyse how people talk about vaccination. Semino and her colleagues explore the language of the pro-vaccination and the anti-vaccination exponents, as well as the undecided population; they believe that the way people talk about this topic mirrors and shapes beliefs, attitudes, and behaviours.

Our study goes in the same direction and it aims to investigate the COVID-19 vaccination debate on Twitter. The language of immunology and virology has been the talk of the town ever since the COVID-19 pandemic started. The effectiveness of official health communication has been

[2] A detailed description of the Quo VaDis project is available at this website: https://www.lancaster.ac.uk/vaccination-discourse/ (8.1.2022).

challenged by a myriad of misinformation, generally spread over the Internet. Zarocostas (2020) uses the term 'infodemic' to define this phenomenon, and a great body of literature has already investigated its impact on the general perception of the COVID-19 pandemic and of the COVID-19 vaccines (Jacobsen and Varga 2020; Garett and Young 2021; Machingaidze and Wiysonge 2021; Kricorian et al. 2021).

In this work, we conduct quantitative and qualitative analyses to examine the COVID-19 vaccine sentiment (at large) and how it is affected in real-time by vaccine news and other information circulating over the Internet. The primary hypothesis advanced by our study is that information on COVID-19 vaccines – in the form of institutional press releases, scientific data, traditional news, and online postings written by social media users – has an immediate effect on the sentiment and the emotions of the general public. This topic is of interest now more than ever, as a negative opinion on COVID-19 vaccines could eventually culminate in vaccine hesitancy. On these grounds, we collect an English corpus of over 214,000 original tweets (over 5.5 million tokens) from March 15th to April 14th, 2021 – a relevant time-frame within the COVID-19 vaccination timeline. Following a multi-method approach, we extract and quantify semantic information from the corpus in the form of sentiment, emotions, and emoji. Concomitantly, the secondary research question of this paper scrutinizes the emoji; we hypothesize that the emoji are able to efficiently evoke concrete and abstract concepts related to the COVID-19 vaccination, and more importantly, they function as emotion enhancers (on either direction of the negative-positive interval), contributing to the sentiment and the emotional dimension of the vaccine discourse on Twitter.

The paper is structured as follows: in section §2 we describe our data and methods; section §3 presents and discusses the results of our analyses; concluding remarks follow in section §4.

2. Methodology

In this section we present our data and methods. First, in §2.1 we illustrate the collection, compilation, and preparation of the corpus, and then in §2.2 we describe the systems used for sentiment, emotion, and emoji analysis.

2.1. Corpus collection and processing

This study explores the semantic dimension of the Twitter debate in English concerning the COVID-19 vaccination campaigns around the world. In order to test our hypothesis regarding the effect of the vaccine news and postings on the users' sentiment and emotions, we collect and analyse a large corpus of tweets for a month, from March 15th to April 14th, 2021. This time-frame is particularly relevant within the COVID-19 vaccination timeline as it covers, among other things, the suspension of the AstraZeneca vaccine[3] in several European countries, in Canada, and in Australia, due to blood clots concerns; the European Medicines Agency (EMA) vaccine review; the discovery of 29 million doses of AstraZeneca vaccine in a Catalent facility in Italy; the rollout and the shipping of Janssen (the official name of the Johnson and Johnson COVID-19 vaccine); the administration of over 150 million vaccine doses in USA; etc.

The data collection process is automatized with the Standard Search Application Programming Interface[4] and the rtweet package (Kearney 2019) for R (R Core Team 2021). In practice, the first step consists in the definition of a list of hashtags that are associated with the COVID-19 vaccination[5]: *#vaccine*, *#vaccines*, *#vaccination*, *#covidvaccine*, *#covidvaccines*, *#covidvaccination*, *#sarscov2vaccine*, *#coronavirusvaccine*, *#coronavirusvaccines*, *#coronavirusvaccination*, *#covid19vaccine*, *#covid19vaccines*, *#covid19vaccination*, *#covid_19vaccine*, *#covid_19vaccines*, *#covid_19vaccination*, *#pfizer*, *#pfizercovidvaccine*, *#pfizerbiontech*, *#pfizervaccine*, *#comirnaty*, *#astrazeneca*, *#astrazenecavaccine,* *#oxfordvaccine*, *#oxfordastrazeneca*, *#vaxzevria*, *#vaxzevriavaccine,* *#vaxzevriaformerlyknownasastrazeneca*, *#moderna*, *#modernavaccine,* *#mrna*, *#mrnavaccine*, *#sputnik*, *#sputnikv*, *#sputnikvaccine*, *#johnsonandjohnson*, *#johnsonvaccine*, *#johnsonandjohnsonvaccine*, *#janssen*, *#janssenvaccine*. Every twelve hours, every day, all tweets (including retweets and quotes) written in English that correspond to these hashtags are automatically downloaded and stored. The data was collected from March 15th to April 14th, 2021 and it amounts to 1,064,936 tweets, corresponding to 31,093,839 tokens. In addition to the text of the tweet, we collect 88 metadata describing the tweet (e.g., character length, number of retweets, number of likes, etc.) and the user (e.g., username, gender, etc.).

In order to reduce the noise in the corpus and to ensure its suitability for linguistic analyses, several processing steps are necessary. First of all, retweets are removed with the *filter()* function available on the dplyr package (Wickham et al. 2020) for R. Besides that, duplicates other than retweets are removed with the *distinct()* function available on the same R package. The final compiled

[3] Until March 25th, 2021 the vaccine was called COVID-19 Vaccine AstraZeneca. After that date the name was changed into Vaxzevria. In this paper we will refer to this vaccine by its former name.

[4] The Standard Search Application Programming Interface is available through the Twitter Developer Platform: https://developer.twitter.com/en (24.8.2021).

[5] On Twitter, the difference between upper and lower-case is not taken into consideration for the retrieval of hashtags, while the "-" character is not supported (the "_" character is used instead).

corpus consists of a data-frame of 214,439 original tweets, corresponding to 5,536,886 tokens[6].

In view of the quantitative analyses, the definition of a 'stop-words' list for English is also necessary. It consists of lexically empty or uninformative words (e.g., prepositions, conjunctions, auxiliary verbs, determiners, etc.), numbers, punctuation, one-character sequences (except for emoji), Twitter handles, URLs, and excessive white spaces. The functions in the tidyverse package (Wickham 2019) are used to apply the 'stop-words' list to the corpus. All hashtags are kept because they contain relevant semantic information; multi-words hashtags graphically separated by capitalized letters are automatically split (e.g., from '#GetVaccinated' to 'get vaccinated') using an R function we created for this purpose. Next, the text of the corpus is converted to lowercase. To use temporal variables for the sentiment analysis, the *created_at* metadatum is divided into date and hour. Since one of the analyses presented here focuses on emoji, for normalization purposes, we replace all skin tones (i.e., light, medium-light, medium, medium-dark, and dark) with the standard yellow colour. Finally, the dataset is stored into a data-frame of 92 columns and 214,439 rows that contains the original tweets, the processed texts, the new temporal information, and 88 metadata regarding tweets and users (e.g., location, number of characters, etc.).

2.2. Methods

This section presents the methods deployed to analyse sentiment (§2.2.1), emotions (§2.2.2), and emojis (§2.2.3) in our corpus of tweets. The characteristics of each system are described in detail, highlighting how the mixed-method approach proposed here allows us to explore the construction of the COVID-19 vaccination debate on Twitter.

2.2.1. Sentiment analysis

Sentiment analysis (also opinion mining) is the point of contention of several fields of theoretical and applied research (e.g., artificial intelligence, computational linguistics, computational social science, cognitive science, natural language processing, text analysis, etc.) and it aims at identifying and measuring opinions and affective states. Feldman (2013: 82) defines sentiment analysis as "the task of finding the opinions of authors about specific entities". To date, there are three known approaches to perform sentiment analysis: lexicon-based, machine learning, and a hybrid combination of the two (see Sharma et al. 2020, for a review).

[6] In the spirit of open science and to enhance collaboration and reproducibility, the final corpus is available as a *.csv* file on the Open Science Framework platform: https://osf.io/ztp4a/?view_only=67988b5786ea46b499febd2062673385 (30.9.2021).

Regardless of the approach, one of the most popular tasks in sentiment analysis is polarity extraction from text (i.e., at word, sentence, and document level), namely the classification of an expressed opinion into positive (i.e., in numerical terms, above 0), negative (i.e., below 0), or neutral (i.e., around 0). Due to its potential, sentiment analysis is studied both in academia and industry, but it is frequently applied in the latter field (e.g., to assess customer feedback). Before the COVID-19 pandemic, few studies have discussed the use of sentiment analysis techniques to explore health communication, let alone health communication on Twitter (see Gohil et al. 2018, for a review). Nevertheless, the outbreak of the coronavirus disease has acted as a catalyst for research on opinion mining. Recent research on this topic has explored both the effectiveness of the institutional communication strategies during the pandemic (Wang et al. 2021), as well as the citizens' reactions to this crisis (Chandra and Krishna 2021). When this paper was drafted, some studies had already analysed the Twitter discourse concerning the COVID-19 vaccination by means of sentiment analysis (Marcec and Likic 2021; Yousefinaghani et al. 2021; Lyu et al. 2021 provide medical and sociological perspectives), but none had a linguistic focus towards the infodemic phenomenon.

In this study we hypothesize that the real-time information (e.g., official and institutional announcements, scientific dissemination, traditional news, online postings, etc.) related to the COVID-19 vaccination circulating on the Internet has an immediate effect on sentiment, therefore on how people perceive the vaccine. To explore this specific semantic dimension of the COVID-19 vaccination debate, we use a lexicon-based system to extract sentiment from a corpus of 214,439 tweets (over 5.5 million tokens). Our lexicon-based approach allows us to track the underlying mechanism of sentiment assignment; therefore, we can easily explore how words contribute to specific sentiment scores in tweets, and more in general how the vaccine perception is built. Unlike other methods of sentiment analysis, a lexicon-based system is practical and immediately available to a wide range of scholars, a crucial element for research reproducibility in linguistics.

Among the many tools available, we use the functions of the syuzhet package (Jockers 2017) for R to extract and analyse sentiment. The approach employed here relies on readily available sentiment lexica able to score the sentiment of a tweet by aggregating the sentiment scores of all the words in the tweet. Generally speaking, these lexica contain words and corresponding sentiment scores ranging from (extremely) negative to (extremely) positive values. Thus, the performance of a lexicon-based approach to sentiment analysis is determined by the fitness of the lexicon.

There are five readily available lexica for sentiment analysis on the syuzhet package, and here we test two of them: Finn's (2011) *afinn* lexicon – created specifically for sentiment analysis tasks in microblogs (e.g., Twitter) and Liu's (2015; Liu et al. 2005) *bing* lexicon – suitable, in general, for opinion mining on the web. The *afinn* lexicon consists of 2,477 words, manually labelled by Finn for sentiment valence (subjectivity, objectivity, arousal, and dominance are not scored). The original score range is comprised between -5 (e.g., 'bastard', 'bitch', etc.) to 5 (e.g., 'hurrah', 'outstanding', etc.); for normalization, comparability, and reproducibility purposes, in this paper the original scores are transformed to match a more common -1 (extremely negative) to 1 (extremely positive) range. Internet slang and acronyms (e.g., 'lol', 'rotflmfao', 'wtf', 'wowow', etc.) are also included in the *afinn* lexicon to better capture the sentiment of the Twitter communication. The *bing* lexicon consists of 6,786 words, classified into two categories: negative (e.g., 'abominable', 'pain', 'scary', etc.) and positive (e.g., 'elegant', 'love', 'smile', etc.), that are transformed in this paper into discrete numeric values (i.e., -1 and 1, respectively). Liu and the colleagues labelled manually only a small list of seed adjectives, by using the 'positive' and the 'negative' tags; this list was automatically enriched and labelled with the support of WordNet (Miller 1995; Fellbaum 1998). To overcome inflection issues, we lemmatize both the tweets and the lexica with the lemmatization functions of the UDPipe package (Wijffels 2021) for R, using the english-ewt (Silveira et al. 2014) pre-trained model.

The extraction of sentiment is performed with the *get_sentiment()* function that iterates over the vector of tweets and assigns two sentiment scores to each tweet, one based on the *afinn* lexicon and the other based on the *bing* lexicon. Two large numeric vectors are obtained corresponding to the two methods. Next, to measure the overall sentiment scores and to ensure comparability across scales and lexica, we apply the *rescale_x_2()* scaling function and the *get_dct_transform()* time normalisation and shape smoothing function. Each tweet, its sentiment score, and the date of publication are stored in a data-frame. In order to obtain a visual representation of sentiment from March 15th to April 14th, 2021, we plot the sentiment scores on a normalised time axis; to do so we apply the *simple_plot()* function to the sentiment vector. This function exploits three smoothing techniques (i.e., rolling average, Loess – local polynomial regression fitting, and Syuzhet DCT – discrete cosine transformation). To explore in detail the effect of real-time news and online postings on the perception of the COVID-19 vaccines, a qualitative analysis is performed. To this end, 99 tweets are extracted from the corpus following a stratified random sampling that controls the time variable (i.e., the date of publication of the tweet) and the sentiment score. The scores are transformed into the corresponding categorical values (i.e., negative, neutral, and positive), resulting into 33 tweets for each label. Some examples pulled from this sample are provided in Annexes A and discussed in §3.1.

The performance of the sentiment analysis system is assessed by three native speakers of British English (one male and two female language teachers working in Bologna). The speakers rate the sentiment of the sample of 99 tweets described above. The inter-rater agreement as well as the agreement between the annotators and the results of the automatic system of sentiment analysis are calculated with Kappa Fleiss test (Fleiss et al. 1969). According to Landis and Koch (1977), Kappa can be interpreted as follows: < 0 = poor agreement, 0.01 – 0.20 = slight agreement, 0.21 – 0.40 = fair agreement, 0.41 – 0.60 = moderate agreement, 0.61 – 0.80 = substantial agreement, and 0.81 – 1.00 = almost perfect agreement. The human annotation enhances both the qualitative and the quantitative analyses.

2.2.2. Emotion analysis

Emotion analysis (also emotion classification or emotion detection) is often seen as a more sophisticated version of sentiment analysis, in the sense that it provides a refined identification of primary emotions in a text (i.e., at word, sentence, and document level). There are three main approaches commonly used in natural language processing to detect emotions: lexicon-based, machine learning, and hybrid systems (see Acheampong et al. 2020, for a review). Unlike sentiment analysis, the emotion analysis does not necessarily employ discrete numeric values, binary variables, or continuous intervals to measure affective states. More commonly, emotions are classified and quantified based on a reference model, generally sourced from psychological research (Combei and Luporini 2021). Accordingly, emotions in text are expressed in terms of levels of categorical variables. The number and the labels of these levels depend on the theoretical model used in the research.

Several theories of basic emotions have been introduced. One of the first examples is James' (1890) model that classifies basic emotions into four categories: *fear*, *grief*, *rage*, and *love*. A hundred years later, Plutchik (1991) proposes an extended list of basic emotions, in the form of a wheel diagram, containing eight emotions: *joy*, *trust*, *fear*, *surprise*, *sadness*, *disgust*, *anger*, and *anticipation*. Based on these emotions, he also hypothesizes the presence of primary, secondary, and tertiary dyads, each containing feelings composed of two basic emotions situated one petal, two petals, and three petals apart, respectively (Plutchik 2001).[7] For example, *remorse* is found in the primary dyad and it is a combination of *disgust* and *sadness*. Among other feelings in the secondary dyad there is, for instance, *hope* which is a combination of

[7] Plutchik (2001) classifies the feelings as follows: in the primary dyad, *love*, *submission*, *awe*, *disapproval*, *remorse*, *contempt*, *aggressiveness*, *optimism*; in the secondary dyad, *envy*, *unbelief*, *despair*, *curiosity*, *guilt*, *hope*, *pride*, *cynicism*; in the tertiary dyad, *anxiety*, *delight*, *sentimentality*, *shame*, *outrage*, *pessimism*, *morbidness*, *dominance*.

anticipation and *trust*. Also, an example from the tertiary dyad is *outrage*, namely a combination of *anger* and *surprise*.

To complement the sentiment analysis, our study exploits a lexicon-based system to detect emotions on Twitter during one month of debate regarding the COVID-19 vaccination. This approach is able to account for the emotional valence of each tweet and to return the most prevalent emotions, both at tweet level and at corpus level. Thus, the users' feelings and reactions concerning vaccination campaigns around the world can be efficiently mapped. The analyses are conducted in R with the syuzhet package, introduced in section §2.2.1.

The lexicon we employ for emotion analysis is called *nrc* and it was created by Mohammad and Turney (2013). This 13,875-words resource is based on Plutchik's (1991) classification of basic emotions and it is the result of a (crowdsourced) manual annotation of emotional valence. Words have an emotional dimension, in the form of one or more basic emotions (i.e., *anger*, *fear*, *anticipation*, *trust*, *surprise*, *sadness*, *joy*, and *disgust*). For example, the word 'agony' is associated with three emotions (i.e., *anger*, *fear*, and *sadness*). Following this approach, words that are not part of the lexicon have no emotional valence for the classification system.

Before detecting the emotions in tweets, we use the tools presented in section §2.2.1 to lemmatize both the corpus and the *nrc* lexicon. We apply the *get_nrc_sentiment()* function to each tweet to extract the most prevalent emotions (a numeric value is provided for each primary emotion existent in the tweet), and then we compute relative percentage-based values for the whole corpus. Emotions are structured and plotted with the functions of the tidyverse and ggplot2 (Wickham 2016) packages for R. A qualitative exploration of the emotion analysis is also performed. First, we calculate the central tendency of each of the eight emotions at the corpus level and then we randomly sample tweets the scores of which are higher than these eight average values, for a total of 80 items (ten for each basic emotion). This returns tweets containing a dominant emotion, namely an emotion the score of which outpoints the scores of the other seven emotions. Some examples extracted from this sample are provided in Annexes B and discussed in §3.1.

The evaluation of the results of emotion analysis is performed following the method described in section §2.2.1, except that in this case, the three native speakers use the *nrc* labels discussed above to tag the emotional dimension of 80 tweets; the examples included in this sample display high emotional values for a dominant emotion.

2.2.3. Emoji analysis

Emoji are small pictographs equipped with predefined names and Unicode tags (i.e., code points) that are used to represent and evoke both abstract and concrete concepts. Emojipedia[8] – the reference website for the official emoji – classifies emoji into eight categories: smileys and people (e.g., worried face - 😟, police officer - 👮, etc.); animals and nature (e.g., turtle - 🐢, water wave - 🌊, etc.); food and drink (e.g., pizza - 🍕, wine glass - 🍷, etc.); activity (e.g., horse racing - 🏇, swimming - 🏊, etc.); travel and places (e.g., airplane departure - 🛫, desert island - 🏝, etc.); objects (e.g., balloon - 🎈, crystal ball - 🔮, etc.); symbols (e.g., ATM sign - 🏧, musical note - 🎵, etc.); and flags (e.g., chequered flag - 🏁, white flag - 🏳, etc.). These pictograms have been part of the Computer Mediated Communication (CMC)[9] for more than two decades, becoming increasingly popular across cultures and among all age groups.

Danesi (2017) advances the claim that the emoji code may be a form of universal language able to solve problems of comprehension. Conversely, Abel (2020: 34) warns against the use of the "myth of universality", suggesting that emoji are "strongly embedded in cultural conditions". On the same note, some scholars have discussed about other types of variation in the emoji use and interpretation, including gender and generational differences (Prada et al. 2018; Herring and Dainas 2018, 2020), but also idiosyncrasy (Hall and Pennington 2013; Dainas and Herring 2021); all these differences may, in fact, lead to faulty interpretations of the communicator's intentions.

Regardless of whether the emoji use is universal or socio-demographically dependent, we know for sure that people have used emoji intensely and for quite some time both on social media (e.g., Facebook, Twitter, etc.) and in private conversations (e.g., iMessage, WhatsApp, etc.). For this reason, the users might not be fully aware of how emoji have shaped the language they speak (Chiusaroli 2017a, 2017b; Kejriwal et al. 2021). The fact that we include emoji in our communication – even though sometimes we do it without much thought – adds significant semantic and pragmatic information to the message. As a matter of fact, emoji may be employed either for mitigation purposes (e.g., a smiling face – 🙂 used with a request) or to better emphasize a written message (e.g., a crying face – 😢 to convey sadness or pain). Recent research has documented several functions of emoji in CMC: enhancing emotions; functioning as rhetorical devices; changing the register

[8] Emojipedia is available online at this website: https://emojipedia.org/ (26.8.2021)

[9] In this paper, the term Computer Mediated Communication (CMC) is used to refer to any for of human communication enabled by means of two or more electronic devices (e.g., computers, mobile phones, tablets, etc.).

and the style of a message; strengthening the illocutionary force of a speech act; and mitigating face-threatening acts (see Bai et al. 2019, for a review). For instance, the experimental study by Weissman and Turner (2018) shows that the wink emoji (i.e., 😉) induces irony, while Cheng (2017) claims that emoji are included more frequently in positive messages.

In this paper, we hypothesise that emoji boost the emotional valence of tweets; together with other semantic features (explored here by means of sentiment and emotion analysis) emoji are able to better reflect the users' perception of COVID-19 vaccination. For this reason, the paper will investigate the semantic contribution of emoji to the sentiment of the Twitter debate on COVID-19 vaccines. Specifically, we will focus on patterns of emoji use in tweets, by measuring their frequencies, examining their concordances, and computing the strength of word-emoji associations. The aim of our emoji investigation is to enhance sentiment and emotion analysis by identifying recurrent features of the COVID-19 vaccine debate on Twitter; a strategic use of semantic polarization and the choice of emotions and emoji may result in persuasive postings that are able to change the users' opinions with respect to vaccination.

From a practical point of view, for the emoji analysis we employ the R packages tm (Feinerer and Hornik 2019), Unicode (Hornik 2020), and emo (Wickham 2020) to process and to analyse our corpus of tweets. In addition to that, we use a 2,455-type emoji dictionary released by Lyons (2017) – based on previous work by Peterka-Bonetta (2015) and the lexicon of emoji sentiment by Novak et al. (2015). These resources allow us to explore the emoji contribution to the sentiment in the corpus, to compute the frequency of each emoji type, and to extract the emoji that are strongly associated with the keywords of the COVID-19 vaccination discourse on Twitter (i.e., "vaccine" "vaccination", and "vaccines")[10].

3. Results and discussion

The main hypothesis of this study is that the abundance of real-time announcements, news, and online postings regarding the COVID-19 vaccination has an immediate impact on the perception of the general public about the vaccines. Since the infodemic relies heavily on manipulative language, a semantic analysis of tweets may be able to uncover clues with respect to the users' sentiment and emotions, which in the long run could contribute to the understanding of the vaccine scepticism. In this section we test this hypothesis in a corpus-based fashion, focusing on the sentiment and

[10] To assess the strength of the association between words and emoji we computed the pointwise mutual information (Ward Church and Hanks 1990).

the emotional valence of the COVID-19 vaccination debate. Moreover, we also verify whether the emoji are able to evoke concepts related to the COVID-19 vaccination, and whether they act as emotion enhancers. In the first part of the section (§3.1), we illustrate the findings of sentiment and emotion analysis, while in §3.2 we describe the results of the emoji analysis; a discussion follows in §3.3.

3.1. The effect of vaccine infodemic on sentiment and emotions

The first results we present and discuss here are the sentiment analysis scores. Table 1 displays the values for central tendency and dispersion at the corpus level. We observe that the overall mean value of the sentiment score in the corpus is just above the neutrality level, reaching slightly positive values, regardless of the lexicon used (0.11 for *afinn* and 0.07 for *bing*). This result, however, is not particularly meaningful for our hypothesis. As a matter of fact, the aggregate sentiment score usually tends to converge toward 0 in case the analysis is conducted on large corpora (Çeliktuğ 2018); in our data this is also indicated by the median values (i.e., 0) obtained with both systems (see Table 1).

Lexicon	Mean	Standard Deviation	Median
afinn	0.11	0.10	0
bing	0.07	0.04	0

Table 1
Sentiment analysis scores.

The large standard deviation suggests dispersion in our results. For this reason and to better depict the users' response to and participation in the vaccine debate on Twitter, we create a temporal representation of the sentiment analysis scores (from March 15th to April 14th, 2021). Since the results obtained with *affin* and *bing* strongly correlate (r = 0.81, p-$value$ < 0.01), and due to page constraints, we include only the plot corresponding to the results of the latter method. Figure 1 shows the scaled sentiment on the y axis, while the time is displayed on the x axis. In addition to the rolling mean (coloured in grey), we include the smoothed curves (Loess in blue and Syuzhet DCT in red).

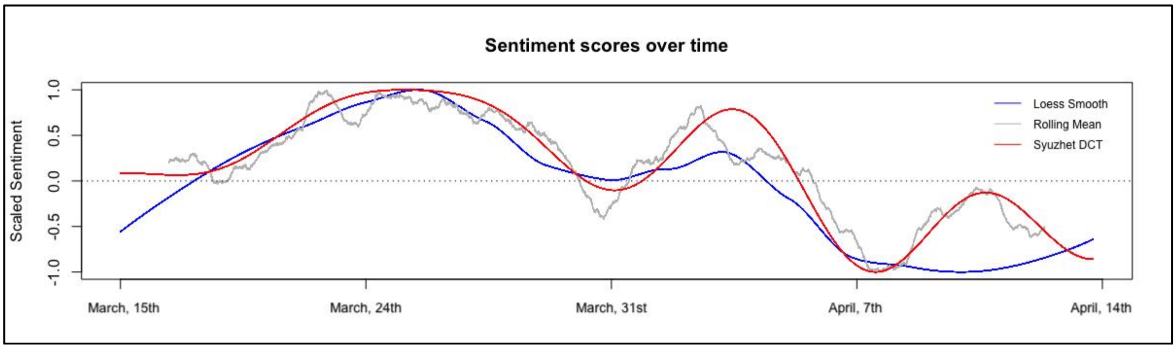

Figure 1
The evolution of sentiment from March 15th to April 14th, 2021.

The graphical representation in Figure 1 allows us to explore the evolution of sentiment. We can observe that during the first days analysed here (from March 15th to March 20th) the sentiment is neutral; the rolling mean is around 0 for most of the time. A qualitative exploration of the tweets written in that period suggests that the score is largely influenced by the decision of France, Germany, Italy, and Spain to suspend the AstraZeneca vaccine over blood clot concerns (see examples 1-3 in Annexes A). There is, in fact, a significant polarization on both sides of the sentiment scale that cancel each other when they are summed up. On the one hand, the positive tweets in our sample refer to messages discussing the vaccine benefits (see examples 4 and 5 in Annexes A). On the other hand, some of the negative tweets suggest that vaccines are dangerous (see examples 6 and 7 in Annexes A), or they reflect the users' concerns regarding the news of side-effects (e.g., fever, blood clots, etc.). There are also several tweets classified as negative that criticize the suspension of the AstraZeneca vaccine; most of these examples are written by British users that blame the EU and EMA for this decision (see example 8 in Annexes A).

The sentiment starts to rise at the end of the first week analysed in our study, in conjunction with the EMA safety review, which outlined the benefits of the vaccines over their side-effects. During the same week, vaccination with AstraZeneca was resumed in most countries. The highest peak of positivity is reached on March 22nd; in our dataset, most positive tweets published on that day are written by users that describe their vaccination experience and that warn against the fake news regarding vaccines (see examples 9-11 in Annexes A). The positivity trend remains relatively stable until March 29th when it starts to fall, reaching a clearly negative score on March 31st. A qualitative exploration of the tweets published at the end of March suggests that the negative score is determined by three key events: German authorities decided to stop the administration of the AstraZeneca vaccine to people younger than 60, following reports of blood clots; Canada suspended the AstraZeneca vaccine shots for people aged 55 and under, as a precautionary measure;

Hungary reported a record number of COVID-19 deaths despite high vaccination rates (see examples 12-14 in Annexes A).

Next, at the beginning of April, the sentiment is neutral and eventually it becomes positive, even if this trend only lasts for a couple of days. Most positive tweets in our dataset refer to the fact that more than 100 million people in the US received at least one dose of a COVID-19 vaccine (with more than 3 million doses administered daily); other positive tweets discuss about the extension of the COVID-19 vaccine to people aged 16 or older in some states in the US (see examples 15-17 in Annexes A). Starting with April 5th, the sentiment falls rapidly, reaching the lowest values on April 7th and April 8th. Despite some fluctuations, the sentiment score remains negative until the 30th day analysed in this study, namely April 14th (see examples 18-20 in Annexes A). Most of the negative tweets in the corpus are reactions to the fact that on April 7th, EMA confirmed a possible link between the AstraZeneca vaccine and events of blood clots; simultaneously, Spanish, Portuguese, and British authorities recommended that younger people should be administered alternative vaccines. Similar decisions were taken in Australia. At the same time, the US Center for Disease Control and Prevention released a statement to address some incidents concerning adverse reactions to the Johnson and Johnson vaccine. The fact that the sentiment score is negative for the entire week suggests that the news regarding the events above (amplified worldwide through Twitter itself) have an immediate effect on the users' confidence in the COVID-19 vaccines. At a more general level, this could be explained in terms of the echo chamber effect, namely a scenario in which perceptions and opinions are magnified and reinforced due to the fact that the communication takes place in a noticeably closed medium. This could also lead to confirmation bias, as the users that look for information regarding the COVID-19 vaccines on Twitter might eventually end up reinforcing their own beliefs on this matter.

The performance of our sentiment analysis system is compared to the performance of three human annotators that rate the sentiment of a stratified random sample of 99 tweets (see also §2.2.1). The results of the first Kappa Fleiss test on sentiment classification suggest a substantial agreement between the three native speakers (tweets = 99, levels = 3, raters = 3, Kappa = 0.777, z = 18.9, p-value < 0.01). We also compute the inter-rater reliability between the human annotators and the automatic classification. The results of this test indicate that the agreement is substantial (tweets = 99, levels = 3, raters = 4, Kappa = 0.798, z = 27.5, p-value < 0.01).

In order to obtain a more detailed perspective of the users' feelings regarding vaccines, but also to better understand how vaccine hesitancy is built as a result of official news and information circulating on the web (including misinformation), we measure the emotional valence of the tweets in our corpus. First of all, a close look at the results spotlights a methodological issue, namely that our system assigns the *trust* label to both the tweets that express trust and to those that express mistrust in the COVID-19 vaccination. For this reason, the plot shown in Figure 2 uses these labels: *trust/mistrust*, *anticipation*, *fear*, *sadness*, *disgust*, *anger*, *surprise*, *joy*.

Figure 2 displays the percentages of all primary emotions in the corpus. We only take into account tweets that display emotional valence, meaning that at least a word in the tweet matches a word in the *nrc* lexicon. The distribution of emotions provides a preliminary response to our research question: the *trust* issues (either *trust* or *mistrust*) regarding the COVID-19 vaccination prevail in our data, amounting to 21.29% of the emotions conveyed; thus, they reflect both trust and scepticism in COVID-19 vaccines. The second most frequent emotion is *anticipation* (16.12%); some tweets display both *trust* and *anticipation* as prevalent emotions (although only one of the two emotions is dominant), generating what Plutchik (2001) defines *hope* – a secondary dyad feeling (see also §2.2.2). The third most frequent emotion in our corpus is *fear* (15.22%); sometimes it occurs together with *anticipation*, an indicator of the users' *anxiety* (a tertiary dyad feeling). Another negative emotion in our corpus is *sadness* (11.8%); when it is combined with *fear* it represents *despair* (a secondary dyad feeling) while with *anticipation* it indicates *pessimism* (a tertiary dyad feeling). The fifth most frequent primary emotion is *joy* (10.93%) that often co-occurs with *anticipation*, implying *optimism* (a primary dyad feeling). It is closely followed by *anger* (10.62%); in some tweets *anger* occurs together with *anticipation*, indicating *aggressiveness* (a primary dyad feeling). Finally, the least frequent emotions in our corpus are *surprise* (8.04%) and *disgust* (5.97%); when they co-occur, they indicate *unbelief* (a secondary dyad feeling).

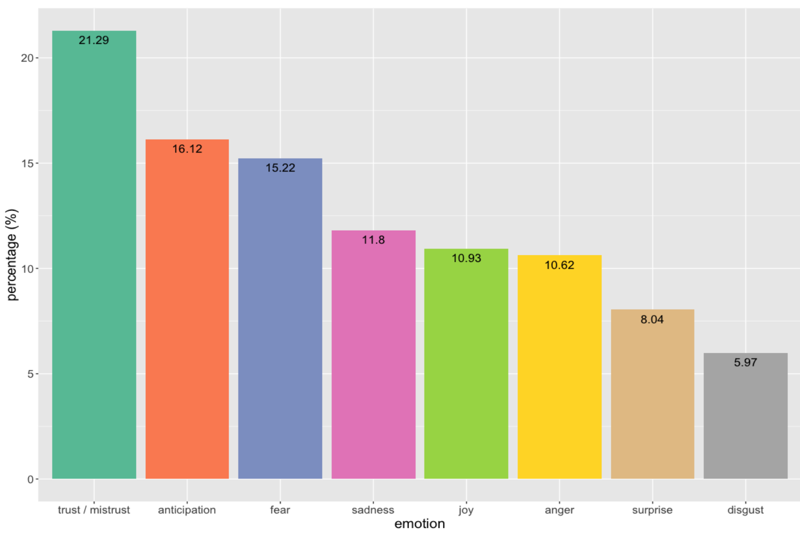

Figure 2
The distribution of emotions.

A close reading of our sample of tweets that convey ‘positive’ emotions (e.g., *trust* and *joy*) and feelings (e.g., *hope* and *optimism*) reveals an interesting finding (see examples 1-5 in Annexes B). Generally, these tweets are written by users that describe positive vaccination experiences and that express gratitude for having received the vaccine (*joy* and *trust* prevail); moreover, several tweets displaying these emotions are written by national and international institutional Twitter accounts that promote COVID-19 vaccination campaigns.

On the other hand, tweets that transmit ‘negative’ emotions (e.g., *mistrust*, *fear*, *sadness*, *anger*, and *disgust*) and feelings (*anxiety*, *despair*, *pessimism*, *aggressiveness*, and *unbelief*) feature a great semantic diversity (see examples 6-10 in Annexes B). Some users express concerns over the vaccine safety, especially in relation to the blood clots incidents reported in Germany and Italy; in most of these cases, *fear* is the prevalent emotion (see examples 7, 10, and 12 in Annexes B). Other users have doubts regarding the vaccine efficacy and effectiveness, probably in response to the news reporting strict confinement measures and significant numbers of COVID-19 deaths in some countries despite high vaccination rates; in these cases, *sadness*, *surprise*, and *anticipation* (and the related feelings of *pessimism* and *disapproval*) prevail (see examples 9, 11, 13, 16, 17, and 19 in Annexes B). Also, some tweets talk about vaccination hesitancy, due to the experimental nature of the vaccines and their accelerated approval; *fear* and *anticipation* (thus also *anxiety*) are frequent (see examples 7, 10, and 17 in Annexes B). There are also tweets that criticize the pharmaceutical industry and, in some cases, the institutions; interestingly, in this case, *anger* and *mistrust* are the dominant emotions (see examples 6, 14, 15, and 18 in Annexes B).

The sample of 99 tweets considered for the qualitative analysis contains several examples of how the COVID-19 vaccine misinformation is built. The purely exploratory analysis of these data reveals an extensive use of negative emotions (e.g., *fear*, *anger*, and *disgust*) and feelings (e.g., *aggressiveness* and *pessimism*) as possible manipulative strategies to amplify the COVID-19 vaccine infodemic. Some users claim that vaccines contain ‘toxic chemicals’ or that they are part of the plan that the pharmaceutical industry and politicians have to ‘inject gene therapy’ or to turn people into ‘robots’ (see examples 6, 8, 14, 15, 17, and 20 in Appendix B).

In order to test the validity of our emotion detection system, its performance is compared to the performance of three human annotators that rate the emotions of a stratified random sample of 80 tweets (see also §2.2.2). The Kappa Fleiss test indicates a substantial agreement between the three human raters (tweets = 80, levels = 8, raters = 3, Kappa = 0.700, z = 28.6, p-value < 0.01). Finally, we calculate the inter-rater agreement between the human annotators and the emotion detection system. These results also suggest that the agreement is substantial (tweets = 80, levels = 8, raters = 4, Kappa = 0.752, z = 43.5, p-value < 0.01).

3.2. The role of emoji

Emoji represent handy resources in the context of Twitter communication since they add relevant semantic and pragmatic information to tweets. In this section we test the secondary hypothesis of this work, namely that emoji are able to evoke both abstract and concrete concepts related to vaccines, and that they enhance the sentiment and the emotional valence of the Twitter debate around the COVID-19 vaccination.

Our analysis focuses primarily on the identification of patterns of emoji use in tweets. A first finding is that our corpus contains 132,203 emoji tokens (with an average of 0.62 emoji per tweet), corresponding to 1,502 emoji types[11]. The type-token ratio is medium-low (i.e., 0.011) and a closer look at the emoji distribution in the corpus suggests that users tend to use few types of emoji very frequently. The twenty most frequent emoji in our corpus are displayed in Table 2. Besides absolute frequencies, we report relative frequencies (per million), for comparability purposes, and the sentiment score associated with each emoji.

Emoji	Absolute frequency	Relative frequency (per million)	Sentiment score (from -1 to 1)
💉	16,960	128,287	0.358
😂	3,623	27,405	0.221
🤔	3,017	22,821	-0.018
🤣	2,674	20,226	0.144
😷	2,661	20,128	-0.169
❤️	2,356	17,821	0.746
🙏	2,254	17,049	0.417
👏	1,969	14,894	0.520
🦠	1,429	10,809	-0.334
😊	1,221	9,236	0.704
😁	1,218	9,213	0.449
🎉	1,212	9,168	0.738
🙄	1,185	8,963	-0.065
💙	1,164	8,805	0.730
💪	1,034	7,821	0.555

[11] According to Emojipedia, as of September 2021, in total there are 3,633 emojis in the Unicode Standard. Further information is available online at this webpage: https://emojipedia.org/faq/ (18.5.2022)

🚨	1,005	7,602	0.638
🥳	901	6,815	0.775
😭	815	6,165	-0.093
😉	798	6,036	0.463
🔥	748	5,658	0.139

Table 2
The 20-most frequent emoji in the corpus.

Unsurprisingly the syringe emoji (💉) is the most frequent in our data; it accounts for 12.83% of the emoji in the corpus. This gives reason to think that in the context of Twitter communication, where characters are limited (i.e., 280), emoji such as the syringe, the face with medical mask (😷), or the microbe (🦠)[12] demonstrate best their iconic and symbolic nature (see examples 1-10 in Annexes C), allowing the users to reiterate the messages and to easily and efficiently represent the desired semantic information.

Positive sentiment prevails in the emoji shown in Table 2. Most of the tweets that contain positive emoji (e.g., 😂, ❤️, 😊, 😁, 🙏, etc.) are written by users that are happy about their vaccine experience (see examples 1, 3, and 11 in Annexes C). Interestingly, in our corpus, the flexed biceps emoji (💪) is used to symbolize the vaccinated arm (see examples 8-10 in Annexes C). Another interesting fact regards the medical mask emoji (😷); according to the lexicon used in this paper, the mask emoji has a negative sentiment. However, in our corpus, it appears in a vast array of tweets, expressing negative, neutral, and positive emotions (see examples 4, 12, and 13 in Annexes C).

The emoji that have a negative valence abound in the corpus, but they are not among the twenty most frequent. Some of the most productive negative emoji are the pouting face (😡), the serious face with symbols covering the mouth (🤬), the flushed face (😳), and the face screaming in fear (😱), that occur mostly in tweets debating the vaccine safety or in misinformation tweets (see examples 14-20 in Annexes C).

The last part of our analysis consists in the extraction of the emoji that are strongly associated with the terms, ‘vaccine’, ‘vaccination’, and ‘vaccines’. Among all the emoji in the corpus, the three keywords are strongly associated with the following emoji types (ordered by the strength of the association): red heart (❤️), check mark (✅), syringe (💉), medical symbol (⚕), flexed biceps (💪), victory hand (✌), male sign (♂), warning (⚠), smiling face with smiling eyes (😊), heart suit (♥), sparkles (✨), double exclamation mark (‼), female sign (♀), face screaming in fear (😱), alarm clock (⏰), skull and crossbones

[12] The high frequency of the microbe emoji (🦠) in the corpus depends on the fact that it reminds of the shape of the coronavirus.

(☠️), high voltage (⚡), question mark (❓), raised fist (✊), coffin (⚰️), hot beverage (☕), exclamation question mark (⁉️), registered (®), frowning face (☹️), and exclamation mark (❗).

Some of the emoji that are strongly associated with the target terms (e.g., ❤️, 💉, 💪, 😊) have already been commented above and they occur frequently in tweets written by people that express gratitude towards doctors, nurses, and institutions, for having received a dose of the COVID-19 vaccine (see examples 1, 2, 3, 4, 6, 7, 8, 10, and 11 in Annexes C). Nevertheless, among the strongest 'vaccin* - emoji' associations we also find emoji that evoke macabre concepts, such as the warning (⚠️), the skull and the crossbones (☠️), the face screaming in fear (😱), the high voltage (⚡), and the coffin (⚰️). A qualitative analysis of the concordances of these associations reveals that in most cases they transmit anti-vaccination messages (see examples 15-25 in Annexes C). This seems to confirm our hypothesis: emoji are indeed able to convey a whole range of concepts linked to the COVID-19 vaccination, both concrete (e.g., the vaccine, the vaccinated arm, the medical mask, etc.) and abstract (e.g., fear, concern, confidence, gratitude, etc.). Furthermore, the results of our analyses indicate that emoji act as stylistic strategies that together with other semantic information (explored here by means of sentiment and emotion analysis) are aimed at supporting and enriching the persuasive and manipulative language of the COVID-19 vaccination infodemic.

3.3. Discussion

The COVID-19 crisis has had a profound impact on public health and it has changed our lives in an unprecedented way. The urgency of the pandemic and the massive investments in pharmaceutical research have contributed to the fast development and approval of several COVID-19 vaccines. At the beginning of 2021, various vaccine campaigns started around the world, prioritising specific groups in the first couple of months, and soon after that, making the vaccines available to the general public. Pharmaceutical companies are now able to produce and deliver vaccines on a large scale; and since vaccines are available, in all probability, a suboptimal vaccination coverage may be caused by vaccine hesitancy. As other scholars have emphasized, this situation may represent a risk for the national healthcare systems, because insufficient vaccination coverage could delay the post-pandemic recovery (Casciani et al. 2021).

Our study contributes to the existing linguistic research on the discourses around the COVID-19 vaccines, by providing new insights on the perceptions and beliefs of the Twitter users. We propose a mixed-method approach that explores the semantic dimension of a large dataset of tweets (over 5.5 million words) written in English from March 15th to April 14th, 2021, by means of corpus-based techniques of sentiment, emotion, and emoji analysis.

One of the first thought-provoking findings of this work concerns the evolution of the sentiment during the month of analysis. The extreme peaks on the time plot and the qualitative analysis of a stratified sample of tweets show that the sentiment score of the Twitter debate on the COVID-19 vaccination is greatly and easily influenced by what is communicated in the media ecosystem such as, for instance, news and comments regarding the decision of several countries to suspend the AstraZeneca vaccine due to blood clots concerns, EMA announcements regarding the review of the COVID-19 vaccines, or reports of record numbers of COVID-19 deaths despite high vaccination rates. Moreover, we observe a significant polarization on both sides of the sentiment scale. These two findings are linked and they are in line with previous research on the topic. As a matter of fact, Jiang et al. (2021) suggest that COVID-19 has become a politicized topic, and the polarization of the debate is a direct consequence of this situation. Furthermore, recent research has demonstrated that Twitter itself encourages the echo chambers effect and the polarization of politicized topics (Cinelli et al. 2021). This happens for primarily two reasons: (1) people choose to follow specific Twitter profiles; and (2) the algorithmic feeds on Twitter are designed to display certain tweets. As a result, users have access mainly to content they already agree with, and their beliefs and perceptions with respect to the COVID-19 vaccines are reinforced or magnified.

The results of the emotion analysis disclose *trust*-related dynamics (either *trust* or *mistrust*) in our corpus. Over 21% of the tweets display *trust* or *mistrust* as dominant emotions and they reflect the users' confidence in vaccines or, on the contrary, the users' vaccine hesitancy. Moreover, since vaccines are perceived as a politicized topic, these emotions refer also to governments and institutions. The fact that *trust* and *mistrust* outmatch the other seven emotions is consistent with recent research on the perception of epidemics on social media. For example, Laurent-Simpson and Lo (2019) claim that there is an overgrowing trend to express mistrust in official public health communication and to discredit institutions. On the same note, Breeze (2021: 10) suggests that mistrust might be "fuelled in many cases by suspicion of 'Big Pharma'". In fact, among the tweets analysed above, there are some clear examples of attacks and criticism towards the pharmaceutical industry. Moreover, *anticipation* and *fear* are well represented in the corpus and they are followed by *sadness* and *joy*; *anger*, *surprise*, and *disgust* (ordered by their frequency) are less frequent. Our qualitative analysis reveals that the infodemic relies on certain negative emotions (i.e., *fear*, *anger*, and *disgust*) and feelings (i.e., *pessimism* and *aggressiveness*) – capable of shaping the users' sentiment regarding the COVID-19 vaccines in the long term.

This study is also complemented with an emoji analysis which shows that emoji represent useful resources on Twitter, since they can evoke both concrete and abstract concepts related to the COVID-19 vaccines (e.g., the vaccine, the virus, fear, gratitude, etc.). Additionally, emoji contribute to the overall emotional content of the Twitter debate regarding the COVID-19 vaccination.

Even though systematic research is required to better understand how the infodemic is constructed in the media ecosystem, our findings suggest that all sorts of vaccine- and health-related information (including dangerous misinformation) – carrying an abundant emotional content – circulate on Twitter and they have an immediate effect on the users' perceptions and beliefs. Misinformation and disinformation represent serious threats for the entire healthcare system, therefore policy makers should develop health communication strategies able to contrast these situations.

After having discussed the results of our analyses, it is important to report the limitations of this study. We will start with the choice of the language. English has an official status in over 60 countries (Adams and Brink 1990) and it is also a global language. However, it is worthy to emphasize that English is a lingua franca only for some Twitter users, typically the most educated. Less educated users or other groups (e.g., disadvantaged people, the elderly, public figures, etc.), but also national institutions, tend to use their native languages on Twitter (Mocanu et al. 2013; Combei and Luporini 2021). Our corpus captures a narrow snapshot of the COVID-19 vaccination debate, as it does not include data from other languages (and consequently other socio-demographic scenarios). A similar limitation is expected as a result of the time variable. While the period considered in this study is pertinent within the COVID-19 vaccination timeline, a month-long corpus collected at the beginning of the global vaccination campaigns can only reflect the Twitter debate during that specific time.

Based on the output of the analyses, the inter-rater agreement tests, and on our own qualitative exploration, we believe that the lexicon-based systems of sentiment analysis and emotion detection proposed in this work are satisfactory. However, since the results depend on the lexica used, up-to-date resources able to reflect the current COVID-19 language on Twitter are needed. The analyses could also be enhanced by means of machine learning or hybrid systems of sentiment and emotion analysis. All in all, additional studies on different languages and time-frames (e.g., later stages of the COVID-19 vaccination campaigns), conducted by means of more advanced techniques would allow us to draw more generalizable conclusions.

4. Conclusions

In the last couple of years, the COVID-19 vaccines have been the predominant topic in the Twitter debate. The paper deployed a multi-method approach to investigate the semantic dimension of this debate, by focusing on the users' affective states, perceptions, and reactions. To this end we collected, compiled, and processed an English corpus of tweets (over 5.5 million words) published from March 15th to April 14th, 2021 – a period that is significant within the COVID-19 vaccination timeline around the globe. We conducted quantitative and qualitative analyses to examine how the perception towards vaccines was altered by news, institutional announcements, and online postings written by Internet users.

Our results showed that the sentiment oscillated during the time-frame considered in this study, displaying polarizations on both the negative and the positive extremities of the continuous sentiment scale. Generally, the positive tweets in this corpus communicated the vaccine benefits and they were written by both institutional accounts and the general public; other positive tweets described the users' personal vaccination experiences and their gratitude towards medical staff. Negative tweets were more semantically diverse expressing, among other things, concerns about the vaccine safety or vaccine scepticism in general, attacks on the pharmaceutical industry, the institutions and politicians, and criticism regarding the suspension of the AstraZeneca vaccine. The main finding of the quantitative and qualitative analyses is that the sentiment of the Twitter users was easily (and instantly) influenced by news, announcements, and online postings regarding COVID-19 vaccination. Presumably, this could reflect the echo chamber effect in the media ecosystem. The beliefs and perceptions of the public opinion with respect to the COVID-19 vaccines were strengthened or magnified due to the fact that the debate took place in a seemingly closed medium.

The fine-grained analysis of emotions performed in this work revealed that the *trust* issues (either *trust* or *mistrust*) outnumbered other primary emotions, amounting to 21.29% of the emotional valence conveyed in the corpus. This mirrors the users' confidence in vaccines or, on the contrary, the users' vaccine scepticism. However, at a more general level, since vaccines have become a politicized topic, this finding could suggest *trust* or *mistrust* in government and institutions. Other recurrent emotions were *anticipation* and *fear*, followed by *sadness* and *joy*; furthermore, *anger*, *surprise*, and *disgust* are less frequent. Interestingly, the qualitative analysis performed on a stratified sample of tweets indicated that the infodemic leant on negative emotions (e.g., *fear*, *anger*, and *disgust*) – able to define and refine the users' perceptions.

Finally, the emoji analysis unveiled that emoji were key resources for Twitter communication. In particular, our analyses showed that emoji were able to evoke both concrete and abstract concepts related to the COVID-19 vaccines. Besides their iconic nature (particularly useful considering the 280-character limit of tweets), emoji functioned as emotion enhancers contributing significantly to the overall sentiment of the Twitter debate regarding the COVID-19 vaccination.

Although additional linguistic and sociological research on this topic is needed, the results of our sentiment, emotion, and emoji analysis seem to indicate that the way the information circulates nowadays in the media ecosystem promotes polarizations with respect to the COVID-19 vaccination – in and of itself a topic capable of being politicized. Therefore, a better understanding of this issue becomes crucial for formulating adequate and inclusive health communication policies and strategies.

Bionote: Claudia Roberta Combei holds a PhD in Linguistics (2019) from the University of Pisa, where she specialised in Phonetics, Corpus Linguistics, and Natural Language Processing. She works as a Researcher (RTD-a) in Linguistics at the University of Pavia where she teaches Phonetics and Phonology, Computational Linguistics, and General Linguistics. Her research deploys mixed-method approaches, by combining computational, quantitative, and qualitative techniques. Some of her recent publications include articles on the semantics and pragmatics of political and media discourse, figurative language, and non-native speech. She has coordinated (with F. Masini and S. Ballarè) the RadioCast-it speech corpus. From 2020, she has been a member of the editorial board of Bibliography of Metaphor and Metonymy (MetBib, John Benjamins Publishing). In 2021, she joined the OED Researchers Advisory Group (Oxford University Press). She is one of the four internal members of CEIVINDICO – an international project funded by UNA Europa.

Author's address: claudiaroberta.combei@unipv.it

References

Abel J.E. 2020, *Not Everyone 💩s Or, the Question of Emoji as 'Universal' Expression*, in Giannoulis E. and Wilde L.R.A. (eds.), *Emoticons, Kaomoji and Emoji*, Routledge, New York, pp. 25-43.

Adams K.L. and Brink D.T. (eds.) 1990, *Perspectives on Official English*, De Gruyter Mouton, Berlin.

Acheampong F.A., Wenyu C. and Nunoo-Mensah C. 2020, *Text-based emotion detection: Advances, challenges and opportunities*, in "Engineering Reports" 2 [7], article e12189, pp. 1-24.

Bai Q., Dan Q., Mu Z. and Yang M. 2019, *A Systematic Review of Emoji: Current Research and Future Perspectives*, in "Frontiers in Psychology" 10(2019), pp. 1-23.

Breeze R., 2021, *Claiming Credibility in Online Comments: Popular Debate Surrounding the COVID-19 Vaccine*, in "Publications" 9, article 34, pp. 1-15.

Casciani F., Pantovic A., Al-Ajlouni Y., Failla G. and Ricciardi W. 2021, *Attitudes, acceptance and hesitancy among the general population worldwide to receive the COVID-19 vaccines and their contributing factors: A systematic review*, in "EClinicalMedicine" 40(2021), article 101113, pp. 1-14.

Çeliktuğ M.F. 2018, *Twitter Sentiment Analysis, 3-Way Classification: Positive, Negative or Neutral*, in "IEEE International Conference on Big Data (Big Data)", Seattle, 2018, pp. 2098-2103.

Chandra R. and Krishna A. 2021, *COVID-19 sentiment analysis via deep learning during the rise of novel cases*, in "PLOS ONE" 16 [8], pp. 1-26.

Cheng L. 2017, *Do I mean what I say and say what I mean? A cross cultural approach to the use of emoticons & emojis in CMC messages*, in "Fonseca: Journal of Communication" 15(2017), pp. 199-217.

Chiusaroli F. 2017a, *Emoticon, emoji e altre logografie per la rappresentazione del parlato nella comunicazione dei social media*, in De Meo A. and Dovetto F.M. (eds.), *La comunicazione parlata / Spoken communication*, Aracne, Roma, pp. 233-251.

Chiusaroli F. 2017b, *Emoji e questioni di genere: una piccola storia della scrittura nella comunicazione digitale*, in De Meo et al. (eds.), *Al femminile: Studi linguistici in onore di Cristina Vallini*, Franco Cesati Editore, Firenze, pp. 137-150.

Cinelli M., De Francisci Morales G., Galeazzi A., Quattrociocchi W. And Starnini M. 2021, *The echo chamber effect on social media*, in "Proceedings of the National Academy of Sciences (PNAS)" 118 [9], article e2023301118, pp. 1-8.

Combei C.R. and Luporini A. 2021, *Sentiment and Emotion Analysis Meet Appraisal: A Corpus Study of Tweets related to the Covid-19 pandemic*, in "RILA [Rassegna italiana di linguistica applicata]" 53 [1-2], pp. 115-136.

Dainas A.R., Herring S.C. 2021, *Interpreting emoji pragmatics*, in Xie et al. (eds.), *Approaches to Internet Pragmatics*, John Benjamins Publishing Company, Amsterdam/Philadelphia, pp. 107-144.

Danesi M. 2017, *The Semiotics of Emoji: the Rise of Visual Language in the Age of the Internet*, Bloomsbury, London.

de Figueiredo A. and Larson H.J. 2021, *Exploratory study of the global intent to accept COVID-19 vaccinations*, in "Communications Medicine" 1 [1], article 30, pp. 1-10.

Feinerer I. and Hornik K. 2019, *tm: Text Mining Package. R package*. https://CRAN.R-project.org/package=tm (21.07.2021).

Feldman R. 2013, *Techniques and Applications for Sentiment Analysis*, in "Communications of the Association for Computing Machinery (ACM)" 56 [4], pp. 82-89.

Fellbaum C. (ed.) 1998, *WordNet: An Electronic Lexical Database*, MIT Press, Cambridge.

Finn A.N. 2011, *A new ANEW: Evaluation of a word list for sentiment analysis in microblogs*, in Rowe M. et al. (eds.), Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages (MSM2011), CEUR-WS, Aachen, pp. 93-98.

Fleiss J.L., Cohen J. and Everitt B. 1969, *Large sample standard errors of kappa and weighted kappa*, in "Psychological Bulletin" 72 [5], pp. 323-327.

Garett R. and Young S.D. 2021, *Online misinformation and vaccine hesitancy*, in "Translational Behavioral Medicine", 11 [9], article 128, pp. 1-2.

Gohil S., Vuik S. and Darzi A. 2018, *Sentiment Analysis of Health Care Tweets: Review of the Methods Used*, in "Journal of Medical Internet Research: Public Health and Surveillance" 4 [2], pp. 1-10.

Hall J.A. and Pennington N. 2013, *Self-monitoring, honesty, and cue use on Facebook: the relationship with user extraversion and conscientiousness*, in "Computers in Human Behaviour" 29 [4], pp. 1556-1564.

Herring S.C. and Dainas A.R. 2018, *Receiver Interpretations of Emoji Functions: A Gender Perspective*, in Wijeratne S. et al (eds.), *Emoji Understanding and Applications in Social Media*, CEUR-WS, Aachen, paper 5, pp. 1-8.

Herring S.C. and Dainas A.R. 2020, *Gender and Age Influences on Interpretation of Emoji Functions*, in "Transaction on Social Computing" 3 [2], pp. 1-26.

Hornik K. 2020, *Unicode: Data and Utilities*. https://cran.r-project.org/web/packages/Unicode/index.html (21.08.2021).

Jiang, J., Ren X. and Ferrara E. *Social Media Polarization and Echo Chambers in the Context of COVID-19: Case Study*, in "JMIRx Med" 2 [3], article e29570, pp. 1-14.

Kricorian K., Civen R. and Equils O. 2021, *COVID-19 vaccine hesitancy: misinformation and perception of vaccine safety*, in "Human Vaccines and Immonotherapeutics" 17 [11], article 1950504, pp. 1-8.

Jacobsen K.H. and Vraga E.K. 2020, *Improving communication about COVID-19 and emerging infectious diseases*, in "European Journal of Clinical Investigation", 50 [5], article e13225, pp. 1-2.

James W. 1890, *The Principles of Psychology*, Henry Holt and Company, New York.

Jockers M. 2017, *syuzhet: Extracts Sentiment and Sentiment-Derived Plot Arcs from Text*. https://cran.r-project.org/web/packages/syuzhet/syuzhet.pdf (20.08.2021).

Kang G.J, Sinclair R.E.N., Mackey L., Schlitt J.T., Marathe A., Abbas K.M. and Swarup S. 2017, *Semantic network analysis of vaccine sentiment in online social media*, in "Vaccine" 35(2017), pp. 3621-3638.

Kearney J. 2019, *rtweet: Collecting Twitter Data. R Package*. https://cran.r-project.org/package=rtweet (15.04.2021).

Kejriwal M., Wang Q., Li H. and Wang L. 2021, *An empirical study of emoji usage on Twitter in linguistic and national contexts*, in "Online Social Networks and Media" 24(2021), pp. 1-23.

Landis J.R. and Koch G.G. 1977, *The Measurement of Observer Agreement for Categorical Data*, in "Biometrics" 33 [1], pp.159.174.

Laurent-Simpson A. and Lo C. 2019, *Risk society online: Zika virus, social media and distrust in the Centers for Disease Control and Prevention*, in "Sociology of Health and Illness" 41(2019), pp. 1270-1288.

Lazarus J.V., Ratzan S.C., Palayew A., Gostin L.O., Larson H.J., Rabin K., Kimball S. and El-Mohandes A. 2021, *A global survey of potential acceptance of a COVID-19 vaccine*, in "Nature Medicine" 27(2021), pp. 225–228.

Liu B. 2015, *Sentiment analysis: Mining opinions, sentiments, and emotions*, Cambridge University Press, Cambridge.

Liu B., Hu M. and Cheng J. 2005, *Opinion Observer: Analyzing and Comparing Opinions on the Web*, in Ellis A. and Hagino T. (eds.), *Proceedings of the 14th international conference on World Wide Web (WWW '05)*, Association for Computing Machinery Digital Library (ACM-DL), Chiba, pp. 342-351.

Lyons K. 2017. *Emoji Dictionary.* https://github.com/lyons7/emojidictionary (21.08.2021).

Lyu H., Wang J., Wu W., Duong V., Zhang X., Dye T.D. and Luo J. 2021, *Social media study of public opinions on potential COVID-19 vaccines: Informing dissent, disparities, and dissemination*, in "Intelligent Medicine" 8(2021), pp. 1-12.

Lyu J.C., Han E.L. and Garving K.L. 2021, *COVID-19 Vaccine-Related Discussion on Twitter: Topic Modeling and Sentiment Analysis*, in "Journal of Medical Internet Research" 23 [6], article 24435, pp. 1-12.

Machingaidze S. and Wiysonge C.S. 2021, *Understanding COVID-19 vaccine hesitancy*, in "Nature Medicine", 27(2021), pp. 1338-1339.

Marcec R. and Likic R. 2021, *Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines*, in "Postgraduate Medical Journal" 1(2021), article 140685.

Miller G.A. 1995, *WordNet: A Lexical Database for English*, in "Communications of the Association for Computing Machinery (ACM)" 38 [11], pp. 39-41.

Mocanu D., Baronchelli A., Perra N., Gonçalves B., Zhang Q. and Vespignani A. 2013, *The Twitter of babel: Mapping world languages through microblogging platforms*, in "PLOS ONE" 8 [4], article e61981, pp. 1-9.

Mohammad S.M. and Turney P.D. 2013, *Crowdsourcing a word-emotion association lexicon*, in "Computational Intelligence" 29 [3], pp. 436-465.

Novak P.K., Smailovic J., Sluban B. and Mozetic I. 2015, *Sentiment of Emojis*, in "PLOS ONE" 10 [12], article e0144296, pp. 1-22.

Peterka-Bonetta J. 2015, *Emoticons decoder for social media sentiment analysis in R.* https://www.r-bloggers.com/2015/10/emoticons-decoder-for-social-media-sentiment-analysis-in-r/ (20.08.2021).

Plutchik R. 1991, *The Emotions*, University Press of America, Lanham.

Plutchik R. 2001, *The Nature of Emotions*, in "American Scientist", 89 [4], pp. 344-350.

Prada M., Rodrigues D.L., Garrido M.V., Lopes D., Cavalheiro B. and Gaspar, R. 2018, *Motives, frequency and attitudes toward emoji and emoticon use*, in "Telematics and Informatics" 35(2018), pp. 1925–1934.

R Core Team. 2021, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna.

Sharma D., Sabharwal M., Goyal V. and Vij M. 2020, *Sentiment Analysis Techniques for Social Media Data: A Review*, in Luhach A.K.et al. (eds.), *First International Conference on Sustainable Technologies for Computational Intelligence. Proceedings of ICTSCI 2019*, Springer, New York, pp. 75-90.

Silveira N., Dozat T., De Marneffe M.C., Bowman S., Connor M., Bauer J. and Manning C.D. 2014, *A Gold Standard Dependency Corpus for English*, in Calzolari N. et al. (eds.), Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), European Language Resources Association (ELRA), Reykjavik, pp. 2897-2904.

Wang Y., Croucher S.M. and Pearson E. 2021, *National Leaders' Usage of Twitter in Response to COVID-19: A Sentiment Analysis*, in "Frontiers in Communication" 6(2021), article 732399.
Ward Church W. and Hanks P. 1990, *Word Association Norms, Mutual Information, and Lexicography*, in "Computational Linguistics" 16 [1], pp. 22-29.
Weissman B. and Tanner D. 2018, *A strong wink between verbal and emoji-based irony: How the brain processes ironic emojis during language comprehension*, in "PLOS ONE" 13 [8], pp. 1-26.
Wickham H. 2016, *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag, New York.
Wickham H. 2019, *Welcome to the tidyverse*, in "Journal of Open Source Software", 4 [43], article 1686, pp. 1-6.
Wickham H., François R., Henry L. and Müller K. 2021, *dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr (21.07.2021).
Wickham H. 2020, *emo: Easily insert emoji into R and RMarkdown*. https://github.com/hadley/emo (20.08.2021).
Wijffels J. 2021, *UDPipe Natural Language Processing – Text Annotation*. https://cran.r-project.org/web/packages/udpipe/vignettes/udpipe-annotation.html (21.07.2021).
Wouters O.J., Shadlen K.C., Salcher-Konrad M., Pollard A.J., Laroson H.J., Teerawattananon Y. and Jit M. 2021, *Challenges in ensuring global access to COVID-19 vaccines: production, affordability, allocation, and deployment*, in "The Lancet", 397(10278), pp. 1023-1034.
Yousefinaghani S., Dara R., Mubareka S. and Papadopoulos A. 2021, *An analysis of COVID-19 vaccine sentiments and opinions on Twitter*, in "International Journal of Infectious Diseases" 108(2021), pp. 256-262.
Zarocostas J. 2020, *How to fight an infodemic*, in "The Lancet" 395(2020), article 10225, p.676.

Annexes A

No.	Tweet	Predicted sentiment
1	#AstraZeneca vaxine is dangerous It contains Polysorbate 80 Which clearly states on Google reasons as to why you SHOULD avoid it Blood clots are a side effect as we can CLEARLY see from 21 Countries suspending it's use. Do your research! #Covid_19 #covid #CovidHoax #WakeUp	**negative**
2	Riled by fears of blood clots Europe's big powers have suspended the use of #AstraZeneca vaccine. Germany, France, Italy and Spain are among them.	**negative**
3	Really pleased to have received my first dose of the #Oxford, #AstraZeneca vaccine this afternoon. A big thank you to Julie, who administered my jab, plus all the team at #Dewsbury Health Centre for their amazing work in getting us all inoculated at such a rapid pace.	**positive**
4	I can't agree more with governments of @MalawiGovt, Poland and Canada, the @WHO, and the @EMA_News on #AstraZeneka #vaccine BENEFITS of the jab clearly outweigh HARMS. Poland blames “media-fuelled panic” for EU countries suspending AstraZeneca vaccine	**neutral**
5	I smiled at this, but sanity is returning. France and Italy are resuming use of the #AstraZeneca vaccine. The European Medicines Agency will release its full findings tomorrow but yesterday confirmed the benefits of the vaccine far outweigh any risks.	**positive**
6	I'm confused...Is the government pushing a rushed, dangerous vaccine on the American people against their will, or is Trump being unfairly treated by not being acknowledged as the hero for single handedly creating this life-saving vaccine?! #foxlogic #vaccine	**negative**
7	I'm so glad I received #PfizerVaccine , no problems! 💉 Stay away from #AstraZeneca 👎 too dangerous. Two Danish patients have brain hemorrhages following AstraZeneca jab	**negative**
8	EU countries are playing politics with #AstraZeneca vaccine as they're still pissed at the UK for Brexit. Their decision to halt usage of the vaccine, even temporarily, will kill far more people than any imagined blood clots from its use.	**negative**
9	Feeling incredibly grateful and privileged to have received my first dose of the #AstraZeneca COVID vaccine today. How far we’ve come in a year! #jabdone	**positive**
10	Number 2 is in my arm. Grateful for scientists in the US and around the world. #vaccine #covid #Pfizer #ÖzlemTüreci #uğurşahin	**positive**
11	Wow! I just got a text telling me about available vaccine appointments in my area. I just signed up for both appointments. Hard to describe the feelings I’m having right now. #forevergrateful #nfa #vaccine #Grateful ❤️⚡💙	**positive**

12	#NSTworld #Canadian experts on Monday recommended halting the use of #AstraZeneca Covid-19 shots for people aged under 55, after a small but rising number of patients abroad suffered blood clots.	**negative**
13	An 80-year-old man suffers adverse event post #CovidVaccine, in coma. Serious #AEFI reported in #Bengaluru. He had no comorbidities. Continues to be on ventilator.	**negative**
14	Sooooo....just thinking.... what if the corovirus mutates again and then starts lethally attacking only those vaccinated?🤔😱 #COVID #CovidVaccine	**negative**
15	Starting May 1st, all Oregonians 16 years and older will be eligible to schedule an appointment for their COVID-19 vaccine. #GetVaccinated #covidvacccine #COVIDvaccine	**neutral**
16	Well, about to hit the 100 million #CovidVaccine mark! Exciting! I can't wait to get my #vaccine shot.	**positive**
17	I was lucky enough to receive my second Oxford #AstraZenaca vaccine yesterday, as an #NHS worker. I hope everyone else gets vaccinated soon. Looking forward to returning to some form of normality...and fun. #coronavirus #COVIDvaccine	**positive**
18	#maharashtralockdown #COVID19 #CovidVaccineScam #CovidVaccine Life saving drugs are available in black market but not in open market! Mockery of words largest pharma industry #Pfizer	**negative**
19	Thank you, I read it. What I don't understand is why under 30s are given a choice of vaccine, if the risk factor is minimal. Other countries have completely banned #AstraZeneca for under 60s. I - and many others - are feeling afraid and bewildered by the mixed messages.	**negative**
20	#CovidVaccine seems just to be a scam! I am hearing so many cases of being tested positive with symptoms even after both dozes! If it doesn't guarantee immunity, is it even eligible to be called a vaccine? Moreover many people getting sick after 1st dose! Its #PR and #Business	**negative**

Annexes B

No.	Tweet	Predicted dominant emotion
1	Thank you very very much good sir! I got mine the 11th and second will be April 8th! I can't wait! #VaccinesWork #CovidVaccine #COVID19 #vaccine	**anticipation**
2	Unabashedly and joyfully liking every tweet I see celebrating an individual's vaccine! Strangers of Twitter, I'm so happy for you! #COVID19 #vaccine	**joy**
3	C'mon folks, let's keep getting vaccinated and then keep on with #HandsFaceSpace afterwards. I was initially reluctant to get the #CovidVaccine but some good quality info on nhsleeds website helped me feel confident. You can read it here	**trust/distrust**
4	What a wonderful day it is, 2nd covid vaccine done 😁❤️ #CovidVaccine #Thankful	**joy**
5	Although she had hesitations, CMH Emergency nurse Jackie Spencer decided to get the COVID-19 #vaccine. ""...I trust in the #science and believe that it is the right thing to do to protect my friends and neighbors,"" she says. Learn when you are eligible	**trust/mistrust**
6	#GreatReset #COVID19 #AstraZeneca #Newworldorder #Controversy #WorldEconomicForum Are you fucking politicians going to beat hitlers kill count? PROBABLY YES! STOP RULING COUNTRIES YOU GODDAMN MURDERERS. ROT IN HELL YOU FUCKING SCUMBAGS	**anger**
7	To #EU. #France and #germany should be prosecuted for #scaremongering and placing lives at risk over the #covid19 #vaccine #AstraZeneca. If people die over this then they (France and Germany) are murderers !	**fear**
8	The brainwashing that I see from people makes me sick at times. To think that you will put toxic chemicals into your body knowing that there's a risk of death or serious side affects and not FDA approved. It's quite sickening #COVID19 #vaccine #AstraZeneca #coronavirus	**disgust**
9	So sad that the vaccine @JoeBiden and @KamalaHarris released under their watch is dangerous. Johnson and Johnson was just halted. The vaccines Trump released are safe. Pfizer and Moderna released under Trumps warp speed is safe. HMM. Makes you think. #vaccine #vaccines	**sadness**
10	'Serious side effects' that were utterly disproven. It's about time people woke up to the strings #BigPharma are pulling in a transparent attempt to undermine #AstraZeneca as they are terrified of the company offering a vaccine to the world at cost. #AstraZenecaVaccine #pfizer	**fear**

11	If you die with a cough or a fever, they will do everything they can to classify it as a COVID death. If you die within hours or days after receiving the COVID vaccine, they will do everything they can to protect Big Pharma. #AstraZeneca #Pfizer #JohnsonAndJohnson #moderna	**sadness**
12	#astrazeneca again i am afraid. i might end up in jail at this rate. #COVID19 #vaccines #bloodclots	**fear**
13	Very sad news from #Georgia. Georgian nurse who went into anaphylactic shock after receiving #AstraZeneca #vaccine dies.	**sadness**
14	Fuck you @who how you advice people to take the shit, #AstraZeneca vaccine is disease, after teasted got extremely pain. If other #Pfizer & #Moderna same it will be disaster for world health. Stop spread the headache #COVID19	**anger**
15	If you think it was only 6 cases that got blood clots from the #JohnsonandJohnson vaccine, you have not learned anything about how the government, big pharma, and media lies to you. They have lied to you this whole time. They are lying to you now.	**trust/mistrust**
16	over half of all adults in #Britain have now been vaccinated with one jab of #AstraZeneca... yet infections are still at about 10,000 daily, more than at this time last year w/o vaccination. What follows from this? Than the #AstraZeneca vaccine doesn't work? #covid	**surprise**
17	It's not just six people that have gotten dangerous blood clots. It's likely many, many more. Think twice before you let Big Pharma inject gene therapy into your body #CovidVaccine	**anticipation**
18	Dear Scotty You can stick your #ageist #vaccine BS up your jaxy maayate. I'm not a unit of profit generation for your #AZ cohort. Anything LNP touches turns to sh1t and I don't trust you. Incompetent, unemployable. #auspol #ScottyFromMarketing #ScottyTheGaslighter #vaccinerollout	**trust/mistrust**
19	Hands up anyone who is surprised that the AZ vaccine has been labelled "safe and effective" by the EU regulatory agency. Nope thought not.. me neither. Were the EU right to halt on such weak evidence?! 🤔 #vaccine #vaccination #AstraZeneca	**surprise**
20	fuck a #vaccine, i'll kick this flu with a 99% survival rate in the ass. don't need the gov turning me into a #robot #JohnsonandJohnson #modernavaccine #AstraZeneca #PfizerVaccine #fraud	**disgust**

Annexes C

No.	Tweet
1	We did it! 💉🦠❤️ #buggeroffcovid #covidvaccine #astrazeneca #numberonedone
2	Had mine today #vaccine #AstraZeneca 💉💪
3	I am vaccinated" 🙏💉 #FirstDose #Astrazeneca #Frontliner
4	Booked in for my first Covid jab 💉😷 #vaccine #vaccination #CovidVaccine #JabToBeatCorona
5	The post 💉 shivers aren't no joke 🥶🥶🥶🥶🥶🥶 - barely made it through the night 😭 #AstraZeneca
6	I got my first COVID vaccine today! YEAH! 💪💉 #COVID19Vaccine #AstraZeneca #FirstDose #Coronavirus #SupportTheNHS #Vaccination #Injections #StaySafe #ThankYouKeyWorkers
7	Got jabbed today! ✅ 💉 🥊 🦠 vaccination for the nation! thanks to Aston Villa and the NHS and volunteers for excellent friendly, organised and smooth system #astrazeneca #vaccination #firstjabstoday #covid_19 #fightback
8	2nd #vaccine in the arm 💪 Thanks to UHSFT
9	Becky's last #ReasonableAdjustment for the #CovidVaccine injection is to ask your doctor 👩⚕ for some numbing cream. 🧴 You put this on your arm 💪 before the injection and helps you not feel the needle. 👍 #WorldHealthDay
10	Over 24hrs since first #Pfizer jab and nothing but a sore arm 💪💉 🦠 #thankyouNHS
11	NHSuk Had my vaccine today. Thank you ❤️ #AstraZeneca #NHSheroes
12	🇪🇺💉😷 Why have several European countries suspended use of the #AstraZeneca #COVID19 vaccine, citing fears of blood clots, even as the EU medicines regulator insists there is no evidence of a link and calls for the jabs to continue?
13	TWAT!! 😡 #Coronavirus #Vaccine #DominicCummings ......🗣💩
14	🤬🤬🤬❗❗ Scary! Dr. #Fauci Wants to Start Vaccinating Little #Babies with the Government's Coronavirus #Vaccine
15	😱Here's 19 Reasons I Won't Be Getting a COVID Vaccination... #Covid #COVID—19 #vaccination #vaccine #GreatReset #NewWorldOrder #MARK #markofthebeast
16	The Vaccine: 😱😱😱😱😱😱😱😱😱😱 😱 — The final solution — 😱 😱😱😱😱😱😱😱😱😱😱 #vaccine #finalsolution
17	😡😡😡😡 Biden continues to talk about the virus mutating but when things go wrong many of us know it will be the #vaccine that is making people sick...dont be fooled people, this was all planned and people are stupid for getting a experimental shot.
18	😡😡😡 Mandatory jabs are forced medical interventions without the patients consent. They are a violation of the Nuremberg Code & Human Rights Law. Experimental Covid Jabs for Care Home Staff to be made Mandatory in UK
19	Scary stuff 😳 #CovidVaccine #COVID19 I'm sure there will be plenty more to come out from the guinea pigs who have already taken the #vaccine 🤦♀ #vaccinated
20	😳😳😳 Even more concerned about getting it! Ugh #bcpoli #cdnpoli #covidbc #COVIDCanada #bchealth @adriandix #AstraZeneca #astrazenecavaccine #bced
21	#astrazeneca is COMING 💉🍆

22	Anyone? #AstraZeneca 💉☠️
23	That's why #Pfizer ☠️ #AstraZeneca ☠️ #Moderna ☠️ Are biggest shits on Earth 🤷 #coronavirus #COVID19 #vaccine #vaccination #WW3 #lockdown
24	Fuck the #vaccine 💉 ☠️
25	🚨☠️⚠️ #ASTRAZENECA💉 PLEADS GUILTY TO #HEALTHCARE⛑️ CRIME [2003] ⏰☎️🧯 📍👀📌