How Do Arab Tweeters Perceive the COVID-19 Pandemic?

Bacem A. Essam[1] **· Muhammad S. Abdo**[2]

Abstract

Language reflects several cognitive variables that are grounded in cognitive linguistics, psycholinguistics and sociolinguistics. This paper examines how Arab populations reacted to the COVID-19 pandemic on Twitter over twelve weeks since the outbreak. We conducted a lexicon-based thematic analysis using corpus tools, and LIWC and applied R language's stylo. The dominant themes that were closely related to coronavirus tweets included the outbreak of the pandemic, metaphysics responses, signs and symptoms in confirmed cases, and conspiracism. The psycholinguistic analysis also showed that tweeters maintained high levels of affective talk, which was loaded with negative emotions and sadness. Also, LIWC's psychological categories of religion and health dominated the Arabic tweets discussing the pandemic situation. In addition, the contaminated counties that captured most of the attention of Arabic tweeters were China, the USA, Italy, Germany, India, and Japan. At the same time, China and the USA were instrumental in evoking conspiracist ideation about spreading COVID-19 to the world.

Keywords COVID-19 · R language · Pandemic · LIWC · Psycholinguistics

Introduction

The coronavirus outbreak (COVID-19) has had a decimating effect across the world since leaving China. The pandemic has left the populations panicked given the recent history of Ebola, Zika, and H1N1 outbreak across many nations. This paper explores the central themes that Arab tweeters discussed related to COVID-19 using a computational content-based approach.

The dynamics of social media channels makes them excellent repositories of linguistic data on current events. Therefore the user-generated content on social media streams constitutes rich data for linguistic and psycholinguistic analysis and reveals much about the

✉ Bacem A. Essam
   literaryartrans@gmail.com; bacemessam@cu.edu.eg

1   Department of Computer Science, Cairo University, Giza, Egypt

2   English Language Resource Center (ELRC), Al-Azhar University, Cairo, Egypt

individual psychological states of users when they attempt to deceive, or to express panic, anger or hatred. The detection of psychological themes expressed in the discourse communicated through virtual media greatly appeals to modern researchers. Linguistic cues have been useful indicators of psychological themes. Duran et al. (2010) studied deception and truthfulness in online conversations. They used Coh–Metrix, unique in that it tracks linguistic features based on cognitive and social factors to evaluate deceptive and truthful conversations that occur within the context of computer-mediated communication. Abdelzaher (2019a, b) uses a lexicon-based approach and corpus tools to trace violence on social networking sites. The linguistic examination of words with expressing violence sheds light on understanding persuasive tactics, public perception of abstract concepts and justification of radicalised actions (Abdelzaher and Essam 2019).

Similarly this study traces the linguistic expressions Arabic tweeters used to express their feelings about, and perceptions of, the COVID-19 pandemic. In addition to the thematic categorisation of the most frequent topics related to the pandemic, it explores the psychological implications of these topics and how they are related to the tweeters' attributions of the causes of the pandemic.

"Review of the Literature" section cites related work and epands upon the research questions of this study. The theoretical background that is primarily based on attribution theory is delineated in "Theoretical Background" section. The proposed method, data collection, and procedure of analysis are explained in "Methodology" section. "Results" section 5 presents themes focused on by Arab tweeters and links these with LIWC-based psychological metrics. "Discussion" section discusses the results with relation to cognitive linguistics, and psychology. The final section draws conclusions and offers recommendations for further research.

Review of the Literature

The user-generated content on social media streams reveals much about the psychology if each tweeter in their attempts to deceive or to express panic. Duran et al. (2010) used Coh–Metrix, a unique tool which it tracks linguistic features based on cognitive and social factors, to evaluate deceptive and truthful conversations that occur within a context of computer-mediated communication.

Within the framework of attribution theory, Roundtree (2018) studied situational crisis communication through hospital tweets on the H1N1 pandemic. The study analysed the organisational and the public use of Twitter during crises to post time-specific storm advice, to share and re-tweet the response efforts of organisations, and to provide post-storm updates. Hospitals tweeted and re-tweeted information from news outlets and opinions on reform from the perspective of experts and physicians, more than reporting real-time statistics and reports on confirmed cases. The study concluded that the use of Twitter during health crises is confined to tweeting too general information and re-tweeted media news. Accordingly the current study hypothesises that general coverage of outbreak and media coverage about the spread of the pandemic are among the most frequent themes discussed by Arab tweeters during this crisis.

In a Zika-focused study Stefanidis et al. (2017) examined the psychological implications of the linguistic features of tweets about Zika. They examined the prominence of linguistic and metalinguistic features in interpreting the behaviour of people during a public health emergency. For that purpose, they collected more than six million tweets from December

2015 to March 2016. Their primary focus was on the geographical distributions of these tweets, the types of users, and the issues which arose as a result of the outbreak. Using spatiotemporal analysis, they discovered that concerns of Twitter users regarding the virus extended from local contexts in South America, the origin, to North America, and then etended to a worldwide focus. Global organisations such as WHO and the Centers for Disease Control and Prevention (CDC) were the most dominant accounts during the outbreak. There seems to have been little trust placed in these organisations during 2020. That is to say; Stefanidis et al. moved beyond thematic analysis to a psychological account of the themes and topics discussed during the health crisis. The current study proposes that fear and distrust will feature among the top psychological categories relevant to these concerns.

Kent et al. (2016) used social media platforms such as Twitter, Facebook, forums, Flickr, YouTube, to collate user-generated responses, adding up to 2.2 million posts, which contained language related to obesity. They investigated the public health risk perceptions of the causal connection between obesity and cancer to tailor health promotion interventions. It was concluded that social media streams could reveal authentic public discourse on health topics, and inform current public health practice. This study hypothesises that the collated tweets can provide insights into the mindset Arab tweeters regarding the causes of COVID-19 pandemic situation.

Fu et al. (2016) concentrate on detecting rumour on social networking sites. To distinguish rumours from general content, they propose a rumour detection method based on implicit features of both contents and users. Such features focus on popularity orientation, internal and external consistency, sentiment polarity and opinion of comments, social influence, re-tweet influence, and the degree of correspondence between messages.

Given that the COVID-19 pandemic has dominated the social media streams, this paper studies representative samples of user-generated content that were tweeted from February 1st 2020 to April 30th 2020.

Theoretical Background

Numerous versions of Attribution theory (Heider 1982; Ross 1977; Weiner, 1986) are concerned with the attempts of ordinary people to understand the causes and implications of the events they witness. Attribution Theory is based upon the premise that people endeavour to know the reasons behind the occurrence of particular events, especially crises. To find causal interpretations for these events, people usually explore the social and the physical world. Besides having a crucial impact on the behaviour of individuals, the perceived causes of a given event result in immediate consequences for manifold psychological processes, including emotional reactivity. Weiner (2010) also established that negative emotions such as anger, envy, and Schadenfreude (joy at the failure of others), and positive emotions such as sympathy and gratitude are triggered by beliefs about causal controllability concerning the behaviour of others.

Causal attributions play a critical role in human behaviour, for they establish an individual's perception of the causal structure of the world and hence serve as determinants of their interactions with that world (Kelley and Michela, 1980). In other words, the way a particular crisis is framed can contribute to a person's appraisal of the cause of that crisis, and thus how they attribute blame (Coombs, 2007). The framing here refers to the cues that are used by stakeholders to make sense of crises or by a psychologically intuitive

person who wishes to explain behaviour and draw inferences about different actors and their contextualisation.

Heider (1958, 1982) differentiates between two kinds of attribution, internal attributions (also referred to as dispositional) and external attributions (also known as situational). People often attribute culpability for negative events to others (external), whereas they praise themselves for positive situations (internal). Kruglanski (1987) argues that the attribution paradigm appertains to the analysis of conspiracy theories as it explains the ascriptions of causality by laypersons, including the negative causality depicted by conspiracy theorists.

Given that language expresses attribution judgments of events manifested by people through distinct language choices, there is a pervasive tendency for people to use more abstract language to describe a high-attribution behaviour, actor, or event. By applying Linguistic Category Modelling (LCM) and attribution theory, Borden and Zhang (2019) linked language tendency and attribution to descriptions of organisations and organisational crisis. More abstract language is used to describe organisations that they perceive to be of enduring problematic internal issues (i.e., high-attribution, preventable crisis) than those they deem to be the victim of a crisis (i.e., low attribution).

From a psychometric perspective, even though function words such as pronouns, conjunctions, and prepositions connote very little content-related meaning, they constitute 55% of word use. In addition to their apparent ubiquity, psychologically speaking style words can be indicative of a writer's personality and psychological traits. Therefore LIWC, frame semantics, and cognitive grammar are the best frameworks for extrapolation from large-scale corpora based on cognitive principles (Pennebaker 1997; Pennebaker et al. 2001; Tausczik and Pennebaker 2010).

Based on the social categorisations of individuals, LIWC organises people's attitudes into four psychological categories. The first of these is 'Analytic,' which comprises the different forms of analytical thinking. This category is a factor-analytically derived dimension that is based on eight function word dimensions. The dimension measures the degree to which individuals use words that indicate logical, formal, and hierarchal patterns of thinking. That is to say, the decline in the value of the analytical thinking measure may relate to individuals who are inclined to reflect their thoughts using narrative language that focuses on the here-and-now, as well as personal experiences. The second category, 'Clout,' highlights the relative social status that reflects the confidence or leadership that people exhibit, whether written or verbally communicated. LIWC's algorithm was designed based on results obtained from a series of studies where people were interacting with one another. 'Tone,' the third category, evaluates the emotional inclination of a writer. LIWC's algorithm for Tone is designed so that the higher the value, the more positive the Tone. Numbers below 50 imply a negative emotional cadence. Finally, 'Authenticity' displays people's perceptions of themselves in a genuine or truthful way. Using a more personal, humble, and vulnerable style was found to be suggestive of Authenticity and honesty. Reference values for every measure are available for comparison on LIWC's website (Cohn et al. 2004; Pennebaker et al. 2003).

This study uses the framework of Attribution Theory to identify and evaluate the causes of the pandemic as perceived by Arab tweeters. It merges the corpus-based semi-supervised approach of analysing and categorising the collated tweets with the findings retrieved from LIWC-2015.

Methodology

We bootstrapped using keywords that are equivalent to "Corona," "Coronavirus," and "COVID-19" pandemic starting from February 1st, 2020 to April 30th, 2020. The final number of tweets totalled 1,920,593 (49,280,161 tokens). A 10,000-tweet sample of the compiled data is appended as supplementary material. We contrasted our corpus against the Arabic Web corpus 2015 (ArTenTen15) to extract unique keywords and n-grams in order to identify characteristic themes in the compiled corpus. Two linguists, acting independently of the aims of the study, annotated the top 1000 keywords and chunks manually according to their respective semantic fields. Their annotations generated the following categories: conspiracy, economy, media, mortality, origin, outbreak and epidemics, politics, prophylaxis, threat, signs and symptoms, and treatment, among others. The differences between their categorisations was statistically insignificant.

We employed the current version of stylo (Eder et al. 2016) to analyse and categorise the posted tweets. Stylo uses an intuitive graphical user interface that parallels would-be written commands. Therefore users of stylo do not need to have an extensive programming background, although familiarity with applied statistics cannot be overemphasised. When the tested documents are uploaded from UTF-8 text files, the most frequent words are automatically extracted and featured as a source for comparing the files under examination.

After analysing the data using corpus tools we applied the LIWC-2015 software program (Pennebaker Conglomerates Inc., Austin, Texas, USA) to retrieve the lexicon-based thematic categorisations for the collected tweets, and to link the generated categories to the overall psychological attitude of the users based on LIWC's algorithm. LIWC contrasts the analysed text with 120 categorising lexica, which are experimentally validated, to provide information on the psychological content of the original writers. These dictionaries were expanded multilingually into Spanish, Dutch, Russian and otherlanguages. We incorporated our version of LIWC-compatible Arabic dictionary in order to categorise words, especially emotion-expressing words. We used the methodological approach of Van Wissen's and Boot (2017) to translate the English version of LIWC-2015 to render our Arabic version. Sketch Engine® and AntConc® were used to generate the concordance of keywords with bilateral 5-word windows for each node. The annotation of themes was manually performed. SPSS 24.0 software for Windows (IBM Inc., Chicago, IL, United States) was used to generate the F-values for one-way ANOVA test, and *posthoc* Tukey HSD, as well as the corresponding statistical level of significance. The p-value of less than 0.05 was considered significant.

Results

Thematic Coverage

Not only has COVID-19 perpetuated fear, political and economic turmoil, global health concerns, and dissent worldwide, the outbreak has suspended human the activity for approximately 15 weeks. On Twitter the dominant themes coronavirus tweets included the outbreak of the pandemic, metaphysics, signs and symptoms in confirmed cases, conspiracism, world and national macroeconomy and microeconomy, politics and

Table 1 Ratios of the themes relevant to the coronavirus until the 12th week after the outbreak

Theme	Coverage	Description	Translated tweets
Outbreak	30.6%	The ways people talked about the outbreak of the virus, new inflicted (confirmed) cases, and global reactions	The global coronavirus outbreak keeps disseminating
Metaphysics	6.8%	Citing verses from scriptures, worship rituals, praying God, inviting people to adhere to moral ethics and religious teachings	We lose lives with this rising death toll and we certainly lose Ramadan
Signs / Symptoms	6.8%	The several earl-stage and late-stage signs and symptoms of coronavirus	The early manifestations of covid19 are not respiratory failure and chills. They can be mild
Economy	6%	Tweets reporting the impact of the virus on stock markets, financial crises, supply chains, working sectors, and working from home	The world_economy and economic growth perspectives decline substantially while we are locked down
Conspiracy	6%	Conspiracist ideation	i think coronavirus is a politic hoax to destroy China
Politics	5.8%	Reactions of governments and political leaders to the outbreak	Trump is making good use of this pandemic
Prophylaxis	4.8%	Tweets providing information about awareness campaigns and ways of protection against the virus, and the materials used for self-protection	Hand sanitization and personal hygiene are the basic fighting tools now
Media	3.8%	Tweets relevant to press releases, breaking news, fake news, myths and news articles."	The WHO news now competes with BBC news
Mortality	3.2%	Death tolls and global death rate	The coronavirus death toll rises as the virus spreads to every country
Threat	3.13%	Fear from the future negative impact of the virus and precautions necessary to resist the panemic transmission	The world is on the verge of the end. All governmental regulation raised the threat level to the critical maximum
Treatment	2.4%	Drugs, vaccines, and steps taken to cure infected people	antimalarial drug chloroquine phosphate antibiotics and zinc are all we have for treating covid19
Sarcasm	2%	This included wordplay and making fun at this gloomy situation	In Egypt, Corona is a piece of chocolate we eat
Origin	2%	Tweets covering places and individuals where the outbreak began	How can we kill those Chinese who ravaged the world with this deadly virus
Strategy	0.73%	Strategies followed by countries to limit the virus outbreak	Call for all frontline medics to provide help at all medical facilities
Research	0.6%	All posts related to efforts dedicated to creating a new vaccine	Vaccine trials for deadly coronavirus is starting with no hope looming
Others	15.34%	They usually include suggestions for continuing education off-campus, the fear of what is coming ahead, the space for sexuality in this new atmosphere and helping the poor. Some tweets reported military acts	How shall college students continue studying during this pandemic?

regulations, prophylactic measures, media and news, mortality, and sarcasm. Table 1 illustrates the ratio and scope of each theme that we manually annotated, along with exemplary tweets.

Outbreak-based tweets were found to be the most dominant over the past twelve weeks. People were particularly concerned about the increasing number of infected cases at the global level. Concerns also to places where the virus is currently spreading. Many users compared the coronavirus to other infectious diseases such as SARS, Bocavirus, Parvovirus, and Pneumonia. The users compared the signs and symptoms of all diseases mentioned above to help each other self-diagnose in the face of the lack of hospitalising medical facilities. Comparisons between coronavirus and 'common flu' or 'common cold' were also carried out across the 12-week period. Further, tweets were full of informative signposts that were apprising people of the signs and symptoms of the novel coronavirus.

Although death-related values might be attributed to the passing news about the mortality rate, speaking about religion gradually increased while leisure-talk and money-speak decreased. A shift towards discussing metaphysical issues was also observed. Muslim tweeters were anxious about missing the holy season of Ramadan and considered the pandemic as divine punishment for the increased committing of sins throughout the world.

Not yet entertaining anxiety about possible starvation, Arab tweeters were anxious about the present and future situation of their national economies and the ensuing financial turmoil. Underlying concerns were directed towards the crashing stock market and excessive returns. Some tweeters predicted an inevitable 'imminent recession'. Countries that received the highest levels of attention were China, the United States of America, Italy, Germany, India, and Japan. Another economic trajectory causing apprehension amongst tweeters was the status was supply chains, with food and logistics being the major topics for discussion. Given the recent decline in the U.S. exchange rate conspiracist ideation concerning hidden agendas were proffered.

Moreover, many linguistic clues highlighted the tendency of Arab tweeters to adopt conspiracist ideation. A few thousand individuals worldwide considered "the whole Corona thing" a scam or a plot directed to sabotage world economies. China also referred to as 'Communist China,' was described as having a 'ruthless' and 'oppressive' regime, which 'caused' the spread of the virus. They also referred to China as the 'black swan,' plaguing the world with human-made lethal contagion. Nonetheless, this conspiratorial attitude became less remarkable by the end of the ninth week after the outbreak. Moreover, other tweets focused on debunking these theories and assuring that the virus is natural. Finally, the virus was also seen as a form of 'propaganda' whose purpose is nothing but 'fear-mongering.'

Since the onset of the outbreak, the different media platforms have played a crucial role in raising public awareness, sharing the latest updates, and providing virus-related information. News covering press conferences, official declarations and responses, and sources of information were the most common. Further, people were encouraged to remain at home. Twitter users were also attentive to 'fake news,' 'disinformation,' and 'hoaxes' that were shared by several 'websites.'

Identifying the first confirmed case in the Chinese city of Wuhan as the origin of the outbreak was considered by Twitter users an invitation to show hostile attitudes towards the citizens of this city as being responsible for the viral infection worldwide. Several accusatory tweets directed towards Wuhan forced the naming of the diagnostic ophthalmologist Li Wenliang. He warned his colleagues about the possibility of facing a global pandemic. Neither the medical institutions nor the Chinese government heeded his warnings, which a

few months later proved to be correct. Turning a blind eye to this warning resulted in the COVID-19 pandemic.

Political figures, governments, healthcare systems, and health officials were the major aspects that the theme of politics encompasses. As for political figures, several (British, Canadian, Japanese) prime ministers were mentioned, the American president and his vice president were also referred to or quoted. As for international governments, Arab tweeters paid due attention to *the Chinese, British, Italian, Australian, and 'local'* governments. Verbs that were used to modify 'government' as a subject included 'handle,' 'announce,' 'stop,' and 'try,' while 'urge' and 'construct' were the main two verbs that were used with 'government' as an object. Healthcare systems were 'overwhelmed,' 'crashed,' and need 'improvement,' according to tweeters. Arabic governments were accused of infidelity and inefficiency in fighting the outbreak.

Prophylaxis included a set of medical and public health precautions necessary to prevent the transmission of the coronavirus. Hand washing, carrying hand sanitiser and wearing disposable face masks are the most frequent means of self-protection. As for the precautionary measurements taken by governments to contain and prevent the further spread of the virus, herd immunity surpassed all other safety measurements in terms of discussion. The imposed quarantine was another subject of much discourse, especially the penalties for those who violate its restrictions. Also included was advice provided by institutional bodies such as WHO and CDC, social distancing as a means of prevention, and finally, hospital-related equipment such as protective gear and testing kits.

Numerical values were most conspicuous with the theme of mortality. China, Italy, India, and Japan received most of the attention. The majority of tweets zoomed in on the rapidly expanding and accelerated pace of new infected cases. Figure 1 shows a word cloud for the incidence of mentioning endemic countries in the studied corpus.

Treatment-based tweets referred to the steps taken by hospitals, healthcare systems, and physicians to treat people who were infected with COVID-19. Doctors and medical staff were lauded for their attempts to raise awareness as regards self-protection. People also questioned drugs being used as vaccines to fight the virus, including antimalaria and antiviral drugs. Many individuals reported their experiences of recovering from a mild corona infection with domestic treatment and bed rest. Nevertheless there was little mention of remdesivir, chloroquine and hydroxychloroquine, ritonavir/lopinavir, as well as ritonavir/lopinavir and interferon-beta.

Coronavirus was also recognised as a threat that could aggravate certain conditions and thus have even more serious repercussions. People compared the virus to climate change, and other global calls which citizens had been warned about, while others criticised people's concern over coronavirus and showed no reaction to violations of government strictures downplaying the gravity of the situation. One subject of tweets which coincided with the theme of conspiracy, is the belief that the world is being prepared for forthcoming biological warfare, and that the virus itself is a biological weapon used by some countries against others.

Governmental strategies were based on recommendations from health organisations. Given that Arab tweeters were suspicious about the credibility of these economic-politically driven decisions, people were sceptical about the travel ban and social distancing. They were specifically concerned about the returning the stranded repatriates who then required obligatory quarantine, and the enforced lockdowns that may have caused deaths not related to COVID-19 (e.g., cardiac and renal patients). However people also referred to

the scientific and experimental efforts exerted with the aim of introducing a vaccine against the virus. Most tweets that addressed this issue focused on news of vaccine trials.

Sarcasm was used to mock the gloomy situation. Tweeters also discussed suggestions for the continuation of education off-campus, the fear lies ahead, reluctance to engaging in sexual activities and helping the poor. Some tweets reported military acts. The percentage rates of all themes discussed, sorted on country of origin for each tweet is shown in Table 2.

Stylo and LIWC

This paper reveals the frequently occurring psychological tendencies inherent in the studied Arabic tweets. Arab tweeters displayed affective psychological processes more than any psychological category. The tweets were also overwhelmed with negative emotions and sadness. And religion appeared among the top five psychological trait. Finally, the analysed tweets typically communicated concerns about the general health issues.

According to the statistical calculations of stylo, the first four weeks (February 2020) were consistent as regards the most frequent words and phrases tweeted, and in turn, the topics chosen for analysis. The fifth and sixth weeks (the ides of March) showed a manifestation of conflicting and diverse topics. During the seventh week the greatest number of tweets containing elements of panic and invitation to all individuals to stay home was detected (Fig. 2).

Psycholinguistic analysis of data using LIWC indicates significant and interesting findings. Levels of analytical thinking have been roughly stable and high over time. This suggests that, during the defined 12-week interval, tweeters were maintaining logical, hierarchal patterns of thought, and cognitive complexity in their tweets, rather than the narrative style. Moreover, the high score in clout represents a high level of confidence; this confidence declined over the weeks but was highly increased at the end of the last two weeks. A low value of Authenticity, as well as low and mid-range scores of Tone were observed over the twelve studied weeks. A less-than-50 Tone relates to negativity. The negativity was
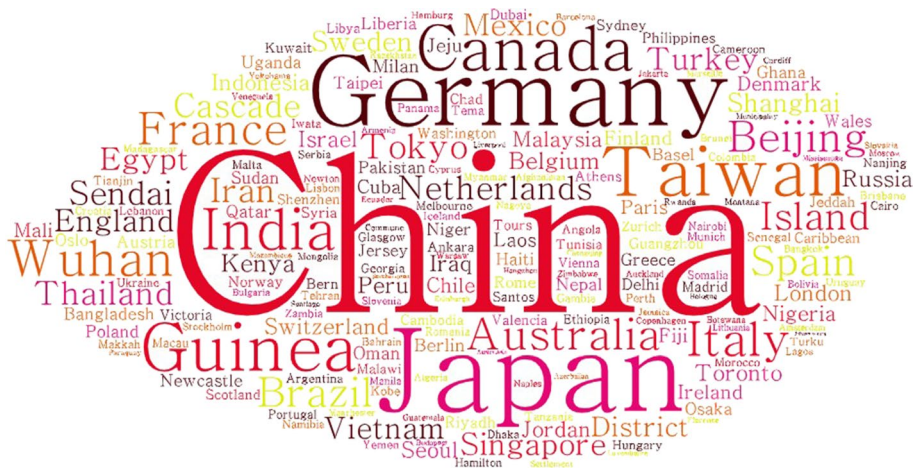

Fig. 1 Word cloud demonstrating the most frequently mentioned countries and cities

suggestive of anxiety and hostility during most of the weeks until a change was observed over the last two weeks of March 2020, when ambivalent feelings began to arise.

In the case of COVID-19, although fear and panic were mitigated by positive approaches to control the pandemic, frustration and dissent began to escalate weekly because of the global organisation and research facilities seemingly playing no effective role in finding treatment modalities if needed treatment options were offered at all.

Psychologically speaking, deception requires more cognitive effort and high processing skills than when telling the truth. Results show that Arab tweeters were both honest and individualised in their communications. This is because high scores of insight-words indicate high levels of individualism and rationality. This is congruent with the mean scores for both Cause and Discrepancy; for, apparently, tweeters were trying to rationalise certain ideas and denying others.

Attribution and Perception

In the causal interpretations that tweeters made in justification of the exponential increases of infections and deaths, many parties were blamed. Arab governments and officials were criticised for not taking necessary and timely precautions to fight the outbreak of the virus. These include economic decisions, precautionary measures, and improper time management. Citizens, too, received constant reproach for not wearing masks, for not abiding by the imposed health curfew, and for not following social distancing guidelines. National and international (Chinese) authorities were accused of not being entirely transparent about the pandemic and for purposefully concealing relevant information.

On a different note, conspiracists attributed the causes of the pandemic to a Chinese plot, whose purpose is to sabotage the word economy. Conversely, others believed it to be an American plot with the aims of crippling Chinese economy. Other conspiracy theories included collusion between the World Health Organization (WHO) and China for manipulating facts about the virus. Finally, some Arab tweeters believed that the virus was created in a laboratory as part of a biological warfare campaign with malicious ecopolitical goals. These intuitive explanations, regardless of their authenticity, aimed to regulate the increased level of anxiety and panic among all citizens.

Discussion

The external cognitive and affective knowledge systems, which are usually based on psycholinguistic or cognitive language resources, operate tools on corpora that are compiled from Twitter. Users of Twitter, as a blog-like social network site, and are constantly applying linguistically unique styles to best communicate their messages within the 140-character limit. However, collecting a representative corpus, with adequate effect size, is inversely proportional to the age of the event the tweets cover, and the intensity of social engagement with a specific event. The greater the number of people who are inclined to interreact positively with a topic over a prolonged period, the less likely collecting a representative corpus from Twitter is achievable. This makes studying a 2-million-token corpus covering approximately the entire time period since the outbreak of coronavirus valuable.

Stylo grouped tweets relevant to similar weeks together. Because the cluster analysis and principal components analysis (with correlation matrix) of the most frequent features

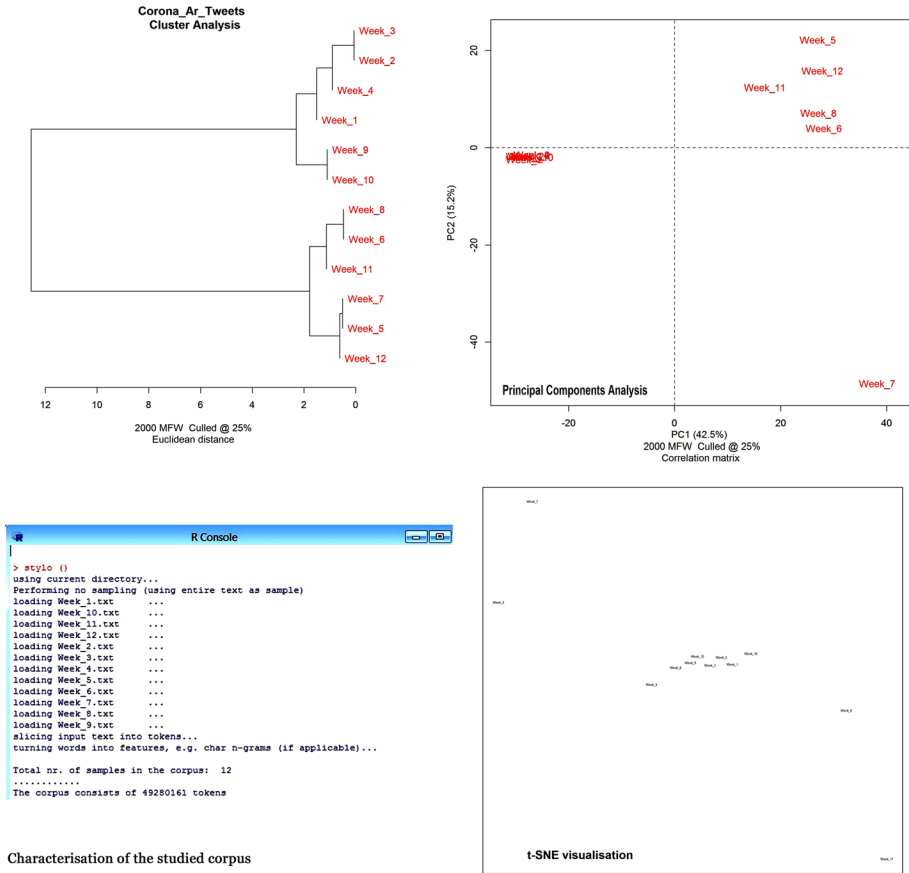**Table 2** The categorisation of the detected themes at the country level

Fig. 2 Similarity between the posted tweets, at a one-week interval, over 12 weeks

render higher accuracy when the tested number of frequent words increase, the number was set at 2000.

We used corpus tools and LIWC-2015 to analyse the compiled corpus. Corpus linguistics has many applications in psycholinguistics (e.g., syntactic priming/persistence), stylistics, sociolinguistics, and forensic linguistics (e.g., authorship attribution) as well as other branches of theoretical linguistics. Two freely accessible corpora were compiled from organisational websites, scholarly articles, preprints and media webpages (Mark Davies's corpus: https://www.english-corpora.org/corona/ and Sketch Engine corpus (https://www.sketchengine.eu/covid19/). We had to compile our corpus from Twitter because we wished to explore public opinion from user-generated content, which is written by Arab populations, as the language of the freely available corpora mentioned above was English.

There is a growing number of corpus linguists who utilise object-oriented programming languages to do their data processing. Given that evaluating the global concerns was our main objective, we did not specifically study the cognitive-linguistic style of each tweeter. Essame et al. (2019) use the principles of corpus-based linguistic cueing, LIWC and FrameNet to explore the common folkloric geopolitical concerns of Egyptian tweeters

about the changing cartography of Egypt. They identified the elements of conspiracy plots and scenarios to conclude the core elements of the proposed deception scenario.

Cohen (2011) used LIWC and the General Inquirer along with corpus tools to conclude that the robust, statistically significant and valence-consistent correlations between the corpus-based dictionaries and measures of psychological distress and dysphoria, have the potential of rekindling the search for theory-driven assessment of emotion dysregulation. Ansumana et al. (2017) explored media hype during the outbreak of the Ebola virus to analyse the misguided conspiracy theories that were propagated by the local media and community gossip about the perceived origin of this pandemic.

The results revealed that community fear, stigma, apprehension, and misunderstanding of the role of healthcare centres operating under the prevailing epidemic conditions encouraged public conspiracist ideation. Similarly, finding a scapegoat to blame for the disaster is consistent with perpetuating global panic so that people will feel that reassured once the culprit has been identified.

Similar to previous reports on facing pandemics (Ferguson et al. 2006), which reported that staying-home prophylaxis coupled with restrictive measures for school closure could reduce transmission rates by 40–50%, the ill-fitting prophylactic measure was deployed to encourage global populations to wear facemasks and disposable latex rubber gloves all day long, even though they are ineffective measures for controlling the infection. People were also discouraged from practising aggregation-based rituals and discretionary travel. Rational and justified as it may seem, this restriction provoked anger among most of the Muslim world. The situation was exacerbated by the beginning of Ramadan. All consolidated organisational effort and published research papers during past global health crises, and the COVID-19 failed to introduce one fast, effective, therapeutic hope to manage COVID-19. So is the growing bulk of contemporary published research, be it original, pseudoscience or replication of previous results. Accordingly, individual attempts of domestic treatment, which were mostly based on folkloric and alternative or herbal medicine, were reported to be the only working solution given the lack of governmental therapeutic intervention plans. The applied governmental retract-and-wait strategies aimed at restricting population movements, prohibiting assembly, and delaying schooling and non-essential work.

Conclusion

To conclude, the study explored the perception of COVID-19 on Twitter, as reflected through Arabic tweets. It, first, identified the most frequent topics discussed in the context of COVID-19 by Arab tweeters. As hypothesised, general news about the outbreak and the number of infected cases ranked highest (30.6%), while circulating medical information about the signs and symptoms of the disease constituted 6.8% of the discourse. Culture-specific talks about religious topics, and metaphysical themes also represented 6.8%. Moreover, relating the situation to hypothetical conspiracies and catastrophic economic consequences were equally among the most invoked topics (6% each).

This paper has also revealed the most frequent psychological categories involved in the Arabic tweets. Arab tweeters expressed affective psychological processes more than any psychological category. The tweets were also pervaded by negative emotions and sadness. Religion appeared amongst the top five subjects for discussion. Finally, the analysed tweets typically communicated concerns about the general theme of health.

This study reveals the causes of the COVID-19 pandemic situation as perceived by Arab tweeters, third. Most Arab tweeters externally attributed the outbreak of COVID-19 to political and economic reasons considered to be part of a Western conspiracy against China. The dominance of these causes corresponded to the thematic prominence of the words "conspiracy", "economy" and "politics". In addition, adherence to metaphysics and religion was encouraged by Arab tweeters to help face the situation. Other more practical, but less frequent, responses included tweeting and re-tweeting prophylaxis and preventive actions.

The tri-phasic analysis portrayed a comprehensive and consistent picture of the perception of COVID-19 by Arabic tweeters. Language is proved to can grant access to the cognition of Arab tweeters and their perception of the crisis and its causes. The most frequent themes determined by the manual annotation of the linguistic content of the tweets were consistent with the automatically retrieved psychological categories of LIWC. Furthermore, the analysis of the concordance lines relevant to the cause-effect relation conformed to the preceding thematic classification and psychological categorisation.

Data Permission: We collected the data for academic purposes (API key: NV77CRHlEkcW9nTNZNOyD****; API secret key: KQQUbwfqIt8ypNpvsuxpIHgDwm Tg70phT93WinFXV4fnCc****).

Compliance with Ethical Standards

Conflict of interest None.

References

Abdelzaher, E. M. (2019a). Lexicon-based detection of violence on social media. *Cognitive Semantics, 5*(1), 32–69.

Abdelzaher, E. M. (2019b). The systematic adaptation of violence contexts in the ISIS discourse: A contrastive corpus-based study. *Corpus Pragmatics, 3*, 173–203.

Abdelzaher, E. M., & Essam, B. A. (2019). Weaponising words: Rhetorical tactics of radicalisation in Western and Arabic countries. *Journal of Language and Politics, 18*(6), 893–914.

Ansumana, R., Keitell, S., Roberts, G. M., Ntoumi, F., Petersen, E., Ippolito, G., et al. (2017). Impact of infectious disease epidemics on tuberculosis diagnostic, management, and prevention services: experiences and lessons from the 2014–2015 Ebola virus disease outbreak in West Africa. *International Journal of Infectious Diseases, 56*, 101–104.

Borden, J., & Zhang, X. A. (2019). Linguistic crisis prediction: An integration of the linguistic category model in crisis communication. *Journal of Language and Social Psychology, 38*(5–6), 650–679.

Cohen, S. J. (2011). Measurement of negativity bias in personal narratives using corpus-based emotion dictionaries. *Journal of Psycholinguistic Research, 40*(2), 119–135.

Coombs, W. T. (2007). Protecting organisation reputations during a crisis: The development and application of situational crisis communication theory. *Corporate Reputation Review, 10*(3), 163–176.

Duran, N. D., Hall, C., McCarthy, P. M., & McNamara, D. S. (2010). The linguistic correlates of conversational deception: Comparing natural language processing technologies. *Applied Psycholinguistics, 31*(3), 439–462.

Eder, M., Rybicki, J., & Kestemont, M. (2016). Stylometry with R: A package for computational text analysis. *R Journal, 8*(1), 107–122.

Essam, B. A., Aref, M. M., & Fouad, F. (2019). When folkloric geopolitical concerns prompt conspiratorial ideation: The case of Egyptian tweeters. *GeoJournal, 84*(1), 121–133.

Ferguson, N. M., Cummings, D. A., Fraser, C., Cajka, J. C., Cooley, P. C., & Burke, D. S. (2006). Strategies for mitigating an influenza pandemic. *Nature, 442*(7101), 448–452.

Fu, K. W., Liang, H., Saroha, N., Tse, Z. T. H., Ip, P., & Fung, I. C. H. (2016). How people react to Zika virus outbreaks on Twitter? A computational content analysis. *American Journal of Infection Control, 44*(12), 1700–1702.

Heider, F. (1982). *The psychology of interpersonal relations*. New York: Psychology Press.

Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process. *Psychological Science, 8*(3), 162–166.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates, 71*(2001).

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*(1), 547–577.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In: *Advances in experimental social psychology* (Vol. 10, pp. 173–220). Amsterdam: Academic Press.

Stefanidis, A., Vraga, E., Lamprianidis, G., Radzikowski, J., Delamater, P. L., Jacobsen, K. H., et al. (2017). Zika in Twitter: Temporal variations of locations, actors, and concepts. *JMIR Public Health and Surveillance, 3*(2), e22.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerised text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24–54.

Van Wissen, L., & Boot, P. (2017). An electronic translation of the LIWC Dictionary into Dutch. In *Electronic lexicography in the 21st century: Proceedings of eLex 2017 conference* (pp. 703–715). Lexical Computing.

Weiner, B. (2010). The development of an attribution-based theory of motivation: A history of ideas. *Educational Psychologist, 45*(1), 28–36.