



CORPORUM JOURNAL OF CORPUS LINGUISTICS

HEC “Y” CATEGORY
ISSN: 2617-2917

Conspiracy Theories (CT) vs Truth Based Reporting: A Corpus Driven Analysis of Covid-19 Online Newspaper(s) Discourse

**Zafar Ullah Shaheen
Ayyaz Qadeer
Fouzia Rehman Khan**

To cite this article: Shaheen et al., (2021) Conspiracy theories (CT) vs truth based reporting: A corpus-linguistic analysis of Covid-19 online newspaper(s) discourse. *Corporum Journal of corpus linguistics*, 4(2), 112-135

**Corpus Research Centre
Faculty of Social Sciences
ISLAMABAD**



Conspiracy theories (CT) vs truth-based reporting: A corpus driven analysis of Covid-19 online newspaper(s) discourse

Zafar Ullah Shaheen

Lecturer in English
Govt. Gordon College Rawalpindi
zafarullahs@gmail.com

Ayyaz Qadeer

Assistant Professor (English)
COMSATS University, Wah Cantt.
ayyazqadeer@yahoo.com

Fouzia Rehman Khan

Assistant Professor, Department of
English SBKWU, Quetta.
fozia.wu@gmail.com

Correspondence:

Zafar Ullah Shaheen
zafarullahs@gmail.com

Received: 18th Dec, 2020

Accepted: 20th Jun, 2021

Published: 23rd Dec, 2021

Keywords:

Chinese, claims, conspiracy, corpus, Covid, theory, post-truth, psychological.

Abstract

In the backdrop of the emergence of conspiracy theories (CT) during the critical days of the pandemic, the discourse of online CT goes unchallenged and has become the part of mundane beliefs. The present study investigates the language/discourse of selected CT online newspaper stories related to the COVID 19 pandemic and compares it with truth-based covid-19 stories. AntConc 3.5.8 (Anthony, 2019) is used as corpus linguistics tool to extract the keywords of the selected newspaper stories, as they are lexical signposts to reveal the most characteristic themes or ‘aboutness’ of the text. A list of keywords generated from the conspiracy corpus includes <China>, <theory>, <conspiracy>, <theories>, <Chinese>, <anti>, <claims>, <wuhan>, <psychological> and <virus>, while the truth-based corpus generated <truth>, <science>, <bullshit>, <posttruth>, <death>, <theory><model> and <covid> as keywords. The keyword list was a handy tool for directing investigators to identify significant lexical differences between both texts and these keywords were further investigated through cluster/N-Grams, concordance and finally for collocates in order to get a more realistic perspective of the keywords generated. Further results showed *conspiracy*, *claims* and *psychological* has co-occurred in the conspiracy corpus and *truth* and *post truth* co-occurred in truth-based corpus.

1. Introduction

The study examines conspiracy theory – (a conspiracy involves a small group of powerful individuals acting in a secret for their own benefit and against a common good) – online discourses taken from different western English language newspapers and compares it with truth based stories. As the spread of covid-19 continues unabated all over the globe, newspapers are full of

stories about its inception, symptoms, care and cure. There are many stories in the newspapers which are based on the facts and truth regarding covid-19 and other stories are based on mere speculations, wishful thinking, rumours and conspirator fantasy. This research paper analyses conspiracy theories related to covid-19 and transports its comparison with truth based covid-19 online newspaper stories. The study is significant among all the information influx regarding the pandemic and search for the contextual clues to comprehend the phenomenon and avoid misunderstanding.

In special circumstances like natural calamities or disasters, conspiracy theories are spread by many corners due to many reasons. Before going in detail of those established reasons it is very important fact that these theories flourish when authentic flow of information is stopped or access to information is restricted; they also get roots when trust on sources and the authorities is low and first-hand information is difficult to come by (Alport & Postman 1981; Fine, 2017; Lewandowsky, 2020; Rosnow, 1980; Shibutani, 1977). In similar circumstances, people exchange their feelings in the form of stories about what they think or believe would be the actual happening; they interchange supposed actions and reactions. Some even extend their imagination to create events as happened before them. Many research studies, as cited above, have been conducted to study the mechanism of the circulation of mostly believable stories in the background of such calamities which have caused sufferings at mega level. Those studies have identified the explanatory role those fabricated stories play at the times of sufferings. (Fine & Turner 2004).

The purpose of these fabricated stories is to make people reach at common point about the causes of that calamity or disruption; it helps to devise certain strategies that are suitable to counter the impending threat, and they may produce certain outcomes, (Tangherlini 20). With a view to study a crisis, stories mostly suggest a specific cause leading to its origin that has generated the threat. Thereafter, they suggest possible solutions or strategies for counteracting in that situation. This pattern of storytelling can be identified in historical events where some calamity has occurred because this pattern tends to repeat itself time and again. For example, the stories of the most significant event of 9/11 had the same pattern as of other events such as the genocidal violence that distorted the face of Rwanda in 1994, and the anti-vaccination attempts at global level (Kitta 2012).

It must be admitted that magnitude of such fabricated and perspectival stories, leading to develop conspiracy theories, is enormous if it is to be collected through traditional means. On the other hand, the proponent of truth based covid-19 stories reject this false reporting and encourage their readers to not pay any heed to these theories and in response to this they have been trying to give opposite view contrary to conspiracy. Now, it has become very difficult for a novice reader to discern through the actual meaning whether to believe conspiracy or truth based news. Apropos, to all this meta-informational discourse, this study will bring about the analytical results of both the newspapers stories in a corpus. It will unearth the most frequent words in the two corpora and the words that are more significant than others within both the texts.

The above discussion can be summed up with a comment that common mind does not take a conspiracy theory as a mere conspiracy theory, but most of it if not all, it is accepted as a fact. They claim and expect that their thoughts and views should be considered as justified since they are full of evidence. However, those facts are convincing only to true believers, and experts. So,

there are two sides of the debate. The conspiracy and truth based reporting and both the narratives are compared statistically in this study in order to determine lexical keywords, frequencies, collocates and concordance. Keywords on its own are not distinctive to carry a meaning unless it is used in a combination with other words (Sinclair, 2005) It would not only inform us of the co-occurrence of related words, but it would also unearth patterns of thought in the forms of keywords of the originators of both the stores.

2. Literature Review

This section delineates the complex concepts of the conspiracy and conspiracy theories as both concepts have been exhausted in the past many years and have been dealt by the researchers and theorists in a befitting manner. A conspiracy comprises of a small group of individuals who are very powerful and are acting in secret for their own vested interests against a common good. The most common conspiracies hatched by the comparatists include Watergate, Iran Contra, and the Tuskegee experiments. The President of the USA, Richard Nixon, along with his administration was allegedly involved in several dubious activities that has undercut the prevailing rule of law and they had sought to punish President Nixon's political opponents for their own gain. Second, Regan administration had carried out Iran Contra scheme that resorts to trade arms to their enemies in exchange for hostages, and surprisingly this was done against the wishes of the Congress. Last, the US government carried out Tuskegee experiment and injected syphilis into the unsuspecting African Americans (Uscinski 2019).

Conspiracy theory explains the past, present, and future circumstances and events that ignite as a primary cause which is a conspiracy. It involves the intentions, planning and actions of powerful people of the world that render them as thoroughly and inherently political. These theories are snappish ideas that may be either true or false, and they are in bare contradiction of the assertion of epistemological authorities which assume such things exist. For instance, claims that assert that George W. Bush administration was behind the attack of 9/11 completely goes against the findings of FBI, CIA, 9/11 commission and other reliable sources (Kean and Hamilton, 2004). Therefore, 9/11 Truth Theories – as they are often called – are suitably called ‘conspiracy theories.’

It is a fact that these rumors and fake stories get set back, but they do not vanish altogether at any stage and they get some ground to have attraction for at least few portion of social fabric, once reliable, authentic and credible information is provided to the masses (Victor, 1993). However, the parallel narratives with their complete frameworks behave like a generative reservoir since side stories never go anywhere. This tradition of interpreting events according to their own aspirations is so strongly footed that, even during normal days when people mostly trust their source of information and have full confidence on the contents of the information spread through credible sources, they don't lose their underlying narrative frameworks – though with low frequency they continue to be spread across numerous social circles. This kind of reservoir of narrative frameworks performs various cultural functions.

There is a strong reason to believe that conspiracy theories, like rumors and legends on which they are based, are opportunistic, taking advantage of low information environments to align the conspiracy theory to unexplained events in the actual news. Such an alignment provides an overarching explanation for otherwise inexplicable events and fits neatly into the world view of the conspiracy theorists. The explanatory power of this storytelling can also entice new members to the group, ultimately getting them to ascribe to the worldview of that group (Bandari et al., 2017). In such conditions there should also be counter narrative which can challenge these prevailing theorists and can put facts and researchers in opposition of them. In the absence of truth based stories, these conspiracies which are spread through conventional and non-conventional media platforms to the larger public, would mislead them for countless number of reason and time.

Social media platforms are used to share online newspaper stories and the information is spread quickly to the larger audience. The present study will look for the informational patterns spread through a network of keywords in online newspapers stories. Concepts in a corpus can be usually revealed by a closer look at keywords and their contexts via concordances (Mahlberg & McIntyre, 2011); keywords and collocates of the keywords and their collocational patterns (Cheng, 2009); keywords in clusters or a sequence of words, bond between keywords (Watson, Todd, 2013); and a combination of these analyses. Further, it will enlighten the readers about the overwhelming choices of words used by the writers to disseminate the conspiracy related hoax and truth based facts.

Given the unsettling nature of data present on the social media and online newspapers regarding Covid-19 pandemic, it is most likely that there could be multiple and conflicting conspiracy theories on these platforms. On the other hand, some of the authors has also given truth based accounts of the pandemic; therefore, it is up to the reader and the researcher to come up with the comparison of the both, and elicit linguistic and non-linguistic patterns in the text. This study is a very first step in the same direction conducted at the outset of the pandemic. It focuses on the linguistic aspect of the pandemic to see how keywords and collocates in a selected corpus are designed in order to influence the minds of the reader and resultantly we have people who would come under the influence of such misrepresentations and misinformation. So, by comparing both conspiracy and truth based stories of the pandemic, we are giving the readers a fair comparison to compel them to decide on their own which side they would sway and go.

3. Methodology

3.1 Introduction to the corpus

This section gives a brief overview of the specialized corpus used for the study. For this article, two different corpora are used by the researchers: one is related to the covid-19 conspiracy theory and the other is related to the truth based covid-19 stories. Different news stories published in the western media regarding the subject of conspiracy and covid 19 are collected, and the total words collected for all the stories were 11264 gathered by remaining within the subject. The other specialised corpus consists of 11227 words on the subject of ‘truth based covid-19’ stories published in different newspapers after the pandemic erupted. The word list of conspiracy related covid-19 corpus derived from the AntConc 3.5.8 is shown in Table. 1 below.

Table 1
Corpus Details

Corpus type	Conspiracy Corpus	Truth-based Corpus
Word Tokens	11264	11227
Word type	2799	2447
TTR	4.024	4.588

3.2 Rationale for choosing the corpus

When corpus is used for the purpose of discourse analysis, it is quite possible to do corpus based analysis on much smaller scale of data instead of millions and billions of words. For instance, Stubbs (1996) has carried out an analysis of two very short letters of Baden Powell that comprised on 330-550 words from each letter. He has successfully shown repetitive differences of how the words were used in his analysis of such a small-scale data.

Moreover, if the researcher is inclined to investigate a specific genre of language and particularly if the genre is limited in some way or the other then it is not mandatory to build a large corpus consisting of million/billions of words. In the same vein, Shalom (1997) has analysed a personal advertisement corpus sent to the London based magazine to get it publish in lonely hearts column. She managed to collect 766 advertisements that made her corpus size close to 15000-20000 words. Apparently, it looks like a very small-scale sample but even then Shalom had demonstrated discernible grammatical and lexical patterns that include frequent occurrence of words like <slim> and <attractive> at the same time.

Lastly, the data that consists of very small scale ‘colony text’ as dubbed by (Hoey 1986) do not necessarily need to be in million/billions of words in length. The average personal advertisement can only be 20-30 words long and each of the advertisement is the embodiment of an individual text and each of the page of it would be called a colony text. In the same manner the order in which these small texts are placed would not change the meaning of the advertisement. Other example of colony text that are present in the newspapers or magazines can include letter to the editors, recipes, horoscope, covid-19 conspiracy theory and truth based stories.

The conspiracy corpus for this study was chosen very carefully by selecting the keywords ‘conspiracy theories about covid 19’ and the search was done through google for the said purpose. There are different stories which emerged, and these were the online version of newspapers published across the globe. All the selected online newspaper stories were taken which were published during the first wave of covid 19 ranging from January, 2020 to June, 2020. In the same way, the other corpus named ‘truth-based’ stories were searched using the keywords ‘truth-based covid 19 stories’ during the same period. In this way search and selection for both the corpus was made for the study.

While building a specialised corpus, instead of the size of the corpus, the fact that how frequent we can find the subject mentioned within the text is kept in mind in order to examine conspiracy and truth in covid-19 online discourse. For instance, while trying to find out conspiracy theories regarding covid-19 in newspapers, two possibilities could be available to choose from. One possibility could be a huge corpus of ten million words or more but mentions conspiracy theories only 20 times; the other possibility could be a corpus which has less words which amount to 20 thousand but mentions conspiracy and covid-19 for 200 times. The first corpus is also useful as a reference corpus which will be discussed in the next section, but the second corpus would tell us more about the conspiracy and covid-19 owing to higher frequency of the subject within the text. In this way we can be more selective when we are going to examine a particular subject while building a specialised corpus on a specific subject. The researcher needs to do this in order to keep the quality of the selected text in particular focus, rather than the quantity of the selected text/discourse. The nature of reference corpus, used in the study as mentioned earlier, is discussed below.

3.3 Reference corpus

The reference corpus is tremendously useful for discourse analysis research and sometimes it becomes necessary to give a detailed brief about it. It is usually very large corpus consisting of millions/billions of words collected from a wide variety of texts that could be the representative of the specific variety of language, but not always linked to the national language. The reference corpus used in order to illicit frequencies and Keynes is also very large – it consists of 741 million words and it is called Coronavirus Corpus. It contains the data on medical, cultural, economic, socio-cultural impact of the coronavirus (covid-19) gathered from online newspapers and magazines from twenty different English-speaking countries (corpusdata.org). The same full-length corpus is used as reference corpus along with covid-19 and conspiracy and truth based corpus.

3.4 Collocational measure

There are three different approaches to calculate collocations including N-Grams, collocation windows and collocation networks. The collocation windows approach is used that goes for co-occurrences within a specified window, i.e. 5L 5R—five words to the left of the selected word and five words to the right, thus giving us the broader associations and pattern of the corpus. The collocation was sorted out by frequency and minimum collocates frequency was chosen as 1 given by default in the AntConc software.

4. Results and Discussion

This section is dedicated to the corpus analysis that includes generating the frequency list of the text, word list and keywords list in order to determine salient features and significant lexical words of the covid-19 text related to the conspiracy and truth based online stories. Frequency list generation is the very first step in such an analysis and goes in the following section.

4.1 Frequency

Frequency list is generated to help the researcher with the important lexical items of a selected corpus. It also helps to investigate the reason why a specific word or word cluster is so frequent in a corpus which ultimately helps to unearth the presence of hegemonic nature of discourses.

Table 2

The fifteen most frequent words (wordlist) in the covid corpus related to conspiracy

Rank	Frequency	Word
1	749	the
2	334	And
3	309	To
4	305	Of
5	281	X
6	246	in
7	234	A
8	166	That
9	133	China
10	124	is
11	116	as
12	99	S
13	95	On
14	93	Covid
15	89	For

The examination of the word lists in Table 2 have given us the picture of all the lexemes used most frequently in the corpus. In the initial Table 1, the lexical words standing out from the whole corpus are definitive article, conjunction, preposition and determiner. All the words are related to grammar and grammatical categories can also hint towards a specific context of the discourse. The other significant words are *covid* and *china* – both are signposts of the emergence of the pandemic itself.

However, in Table 3 below, the most frequent lexical words are very interesting – these words are expected to come out in the same manner. These words are related to the emergence of pandemic, its origin, and consequently spread of the same for instance <china>, <covid>, <virus>, <pandemic>, <world>. There are other words related to the topic of the understudied article <conspiracy> (appears 77 times), <theories> (appears for 58 times), and <international> (appears for 22 times).

In the next step conspiracy theory covid corpus is compared (Table 4) with truth based covid corpus for the purpose of having a fair calculation of the frequencies of both.

Table 3*The fifteen most frequent lexical items (words) in the covid-19 corpus (conspiracy)*

Rank	Frequency	Word
1	749	The
2	281	X
3	166	That
4	133	China
5	116	As
6	93	covid
7	87	By
8	77	Conspiracy
9	59	Virus
10	58	Theories
11	51	Chinese
12	40	Us
13	39	Pandemic
14	37	World
15	32	such

Table 4*Conspiracy theory Covid-19 stories*

Rank	frequency	Word	frequency	Word
1	749	The	371	And
2	281	X	302	X
3	166	That	235	That
4	133	China	162	Is
5	116	As	116	Are
6	93	Covid	111	It
7	87	By	108	We
8	77	conspiracy	93	Covid
9	59	virus	74	not
10	58	theories	68	or
11	51	Chinese	55	there
12	40	us	51	pandemic
13	39	pandemic	41	death
14	37	world	41	science
15	32	such	41	What

As it was hoped that the above table would reveal more interesting differences between the conspiracy related stories and truth based covid 19 stories. Apart from three lexical words which are similar in both the tables, the rest of all lexical items are different: i.e. <covid>, and <pandemic> both are noun and simultaneously present in both the frequency tables, whereas one

word <that> which is determiner/pronoun/conjunction is also present in both tables (it has been used in all forms). Frequency comparison analysis shows that whatever the circumstances are, whether it is a conspiracy or it is truth based online news stories, the two nominal words remain unchanged, i.e. <coved> and <pandemic>. It means that coved and pandemic are omnipresent in all the discourses, and all other dynamics can change, but we can surely say that there is *pandemic* hit which is being reflected in online news discourse and it is labelled as *coved*, and it is both conspiracy and truth in today's global world.

Now let's compare the differences pointed out in the frequency table and see the important lexical words and features of the coved pandemic conspiracy and truth related stories. In conspiracy related list, the pivotal words are <conspiracy>, <virus>, <Chinese> and <world>, while in truth based list <deaths> and <science> are the most important lexical words. It has become clearer by the comparison that the conspiracy related online stories are hit by the same word and on the top of *conspiracy* there is *virus*, and a country name that is *china*, and the aftermaths are being faced by the whole *world* (see Table 4 above). While in truth based list there are two naked truths in the forms of lexical words which are *death* and *science*. The word *death* has got so much importance that cannot be ignored because to date, there are 1.5 million deaths around the globe and the truth based frequency list speaks volume about it. As far as *science* is concerned there is another truth that the solution lies within the word of science and there is no doubt in it. It is because of science we have achieved a milestone of a coronavirus vaccine within the short span of one year and truth based online corpus rightly pointed it out.

4.2 Introducing Keyness

Keyness is another important feature in the corpus analysis; in this feature frequencies are compared with one another in order to find out which word occurs more often statistically than the other word and the list which is generated as the outcome of such practice is called a keyword list. Having such a list gives the researcher the room that can be more useful in suggesting lexical words that need further investigation. Keywords are words which occur with significantly higher frequency in one corpus when compared to some appropriate normative corpus (Scott & Tribble, 2006). Usually, a keyword list is the measure of *salient* feature of the text in comparison with a simple word list that only show frequencies of the given text. Log-Likelihood method is used for the distributional difference while calculating Keyness of the corpus.

Table 5 gives snapshots of the whole covid-19 conspiracy related text and shows all the significant keywords in the list. The top three words are directly related to the pandemic, its place of eruption and the subject under discussion, i.e. <conspiracy>, <theories> and <china>. The other pivotal keywords are <origin>, <wuhan>, <lab>, <psychological> and <virus>. Moreover, the last words that need attention are *lab* and *virus* and both these words signal towards the exact conspiracy which is being talked about a lab in *wuhan* where such a virus was reportedly built. The keyword list has confirmed all the important words signaling towards the pandemic whether it is the country *china*, its city *Wuhan*, where the *virus* got originated even the word *virus* and *origin* are confirming this trend. In addition to the *lab* where it was going to be produced, the last word that needs attention is *psychological* which refers towards the psychological impact of covid-19 pandemic on people and it seems to be severe.

Table 5*Showing keyword list of the covid-19 conspiracy corpus*

Rank	Keyness	Keyword
1	+2650.82	X
2	+682.62	Conspiracy
3	+518.08	Theories
4	+428.76	China
5	+173.88	Ccp
6	+172.84	Chinese
7	+116.76	Theory
8	+107.62	Origin
9	+95.2	Anti
10	+87.89	Claims
11	+86.46	Wuhan
12	+84.12	Psychological
13	+80.08	Lab
14	+76.04	Romanoff
15	+67.56	Virus

In the next step, this keyword list will be compared with the keyword list of truth based online corpus (Table 6) in order to find out similarities and differences in the text.

Table 6*Showing keyword list comparison of the covid-19 conspiracy & truth based corpus*

Rank	Keynes	Keyword	Keynes	Keyword
1	+2650.82	x	+2873.2	x
2	+682.62	conspiracy	+255.89	truth
3	+518.08	theories	+204.84	science
4	+428.76	china	+129	bullshit
5	+173.88	ccp	+122.32	post truth
6	+172.84	Chinese	+121.96	excess
7	+116.76	theory	+104.18	that
8	+107.62	origin	+86.21	deaths
9	+95.2	anti	+77.01	theory
10	+87.89	claims	+72.11	models
11	+86.46	Wuhan	+68.32	nchs
12	+84.12	psychological	+67.95	don
13	+80.08	lab	+66	data
14	+76.04	Romanoff	+64.45	covid
15	+67.56	virus	+63.18	Lauren

The elicited keyword lists of both the corpora related to covid-19 are present in the above displayed table. Interestingly, there is only one similar word that has appeared in both the lists, i.e. <theory>. This word is present in conspiracy theory corpus as plural as well as in singular form, whereas in truth based corpus it appeared only as singular. The rest of the conspiracy list is exhaustive – it includes <china>, <origin>, <anti>, <claims>, <Wuhan>, <psychological>, <lab> and <virus>. All the lexical words in this list have been discussed in the above paragraph to establish all the links with the covid-19 pandemic. The successive discussion is about truth based keyword list focusing on all the differences in the list, starting from <truth>, <science>, <post truth>, <excess>, <deaths>, <models>, <data>, and <coved>.

It is interesting to see that in the truth based keyword, the word that comes on the top of it is <truth> which is self-evident followed by <science>; here we can deduce that all the truth related corpus talks about truth and science and all the related logic of the pandemic. *Truth* is present in the truth based corpus in the same way as *conspiracy* was present in the conspiracy related corpus. The third most frequent keyword in the list is *post truth* which again is very interesting and refers towards the beliefs and emotions of the people around the world towards pandemic rather than believing on the facts and figures. For example, post truth age is characterized by the lack of confidence in the institutions, lack of agreement on facts & figures and having a thin line between facts and opinions. Such beliefs and opinions are shaped by the transcendental perspectives of the people rather than listening to the concrete world of facts.

Moving on, we have two keywords signaling towards the death and excessive deaths that has happened so far since the inception of the pandemic. Whether we believe that pandemic exists or not, one thing is for certain and that hints towards the death toll of all the victims. The last three keywords speak about the models of data which has been created in order to access the colossal damage of the pandemic. For instance, WHO (World Health Organisation), FDA (Food and Drug Administration), EMA (European Medicine Agency) are doing collaborative work in order to have a fair analysis of pandemic happenings, and of developing and approving its cure.

4.3 Analysis of Keywords

The keywords tell us about the ‘aboutness’ of the text in both lists. If we discard all the nouns, proper nouns, conjunctions and prepositions in the keywords’ comparison table (Table 6), there are only few words that need attention after that. First, we would talk about the conspiracy related corpus and the very first word in this corpus is ‘conspiracy’; here we need to know more about this word and without digging into its context we cannot analyze what details it has brought for us being at the top of keyword list. For this, we have to examine all the individual keywords which are strongest, and which appear unusually in the text by carrying out concordances of them.

After performing the concordance on the corpus data, it was found out that the common phrase containing the word *conspiracy* includes *conspiracy beliefs* at the top, *conspiracy believers*, *conspiracy coronavirus* and the rest of them are *conspiracy theories* present in the whole remaining list. Second, most important word is *claim* that appears 21 times in the conspiracy corpus list and after performing concordance operation on the word *claims* it is evident that *claim* is carried with all sorts of prepositions in the list. The third and final word is psychological that starts from *psychological factors*, *psychological impacts*, *psychological needs*, *psychological research* and all the rest is *psychological science*.

The analysis of Cluster/N-Grams (Table 7) and collocates (Table 8), and image of concordance (Fig. 1) below confirm that *conspiracy theories* and *theories* are the most unusually frequent words used in the conspiracy corpus. In addition to this, *conspiracy beliefs* and *covid* are also in the list giving the contextual information that people have confirmed belief in conspiracy theories regarding the covid-19 pandemic. So, whenever conspiracy theories come, they are definitely coming from a certain source and that source remains *conspiracy theorist/s*. The co-occurrence of *conspiracy theory/theorists*, *beliefs* and *covid* are statistically determined through the thorough examination of the listed techniques.

Table 7

Clusters/N-Grams of the keyword conspiracy

Rank	Frequency	Clusters
1	49	conspiracy theories
2	12	conspiracy theory
3	05	conspiracy theorists
4	04	conspiracy theorist
5	02	conspiracy beliefs

Clusters/N-Grams are used to summarize the results generated by concordance tool and it also shows some of the common expressions in the corpus. In the conspiracy theory corpus, the phrases *conspiracy theories*, *conspiracy theory* and *conspiracy theorists* remained at the top of the list. The result is clear to the extent that conspiracy occurred with theory(s) with the further help of theorists. Conspiracy in the case of *covid* is attached with conspiracy creators. In the UK, people who allegedly believe that 5G cellular networks are the cause of COVID-19 set fire to multiple cell towers (Slotkin, 2020)

Table 8

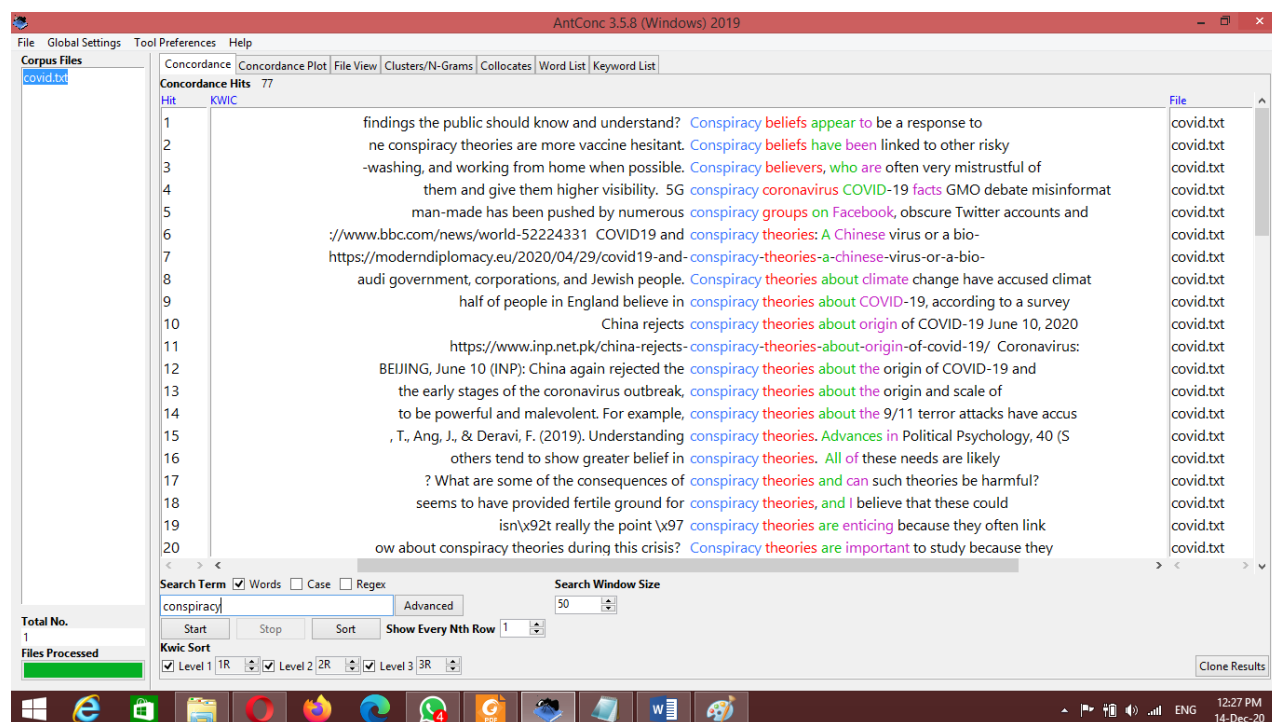
Collocates of the keyword conspiracy

Rank	Frequency	Collocates
1	62	Theories
2	24	The
3	23	To
4	20	Of
5	19	Covid

Collocates tool allows us to see the non-sequential patterns of language in a corpus of text. The representation of the conspiracy collocates is done by dividing the results into three categories of meaning: *theories*, *the*, *to*, *of*, and *covid*. The colligation pattern of (the, to, of) is grouping of words that is based on the way they function in a sentence structure. We can observe that collocation is on lexical level while colligation is on syntactic level.

Figure 1

The concordance chart of the keyword conspiracy



Concordance tool is used to show results of the keywords in a context, and it further tells us how words and phrases are commonly used in a corpus of text. Concordance of the keyword *conspiracy* elicited *conspiracy beliefs* at the top, *conspiracy believers*, *conspiracy coronavirus* and the rest of them are *conspiracy theories*. COVID-19 conspiracy beliefs are strongly related to several psychological and political factors (Uscinski et al. 2020).

Table 9

Cluster/ N-Grams of the keyword Claims

Rank	Frequency	Clusters
1	03	Claims that
2	02	Claims by
3	02	Claims have
4	01	Claims without
5	01	Claims about

The clusters/N-Grams result is not much impressive and only elicited three words out of which one is connector/determiner, second is preposition and third is an auxiliary verb. When party leaders and media personalities promote conspiracy theories and misinformation, likeminded individuals exposed to this rhetoric are more likely to adopt those ideas (Swire et al., 2017).

Table 10

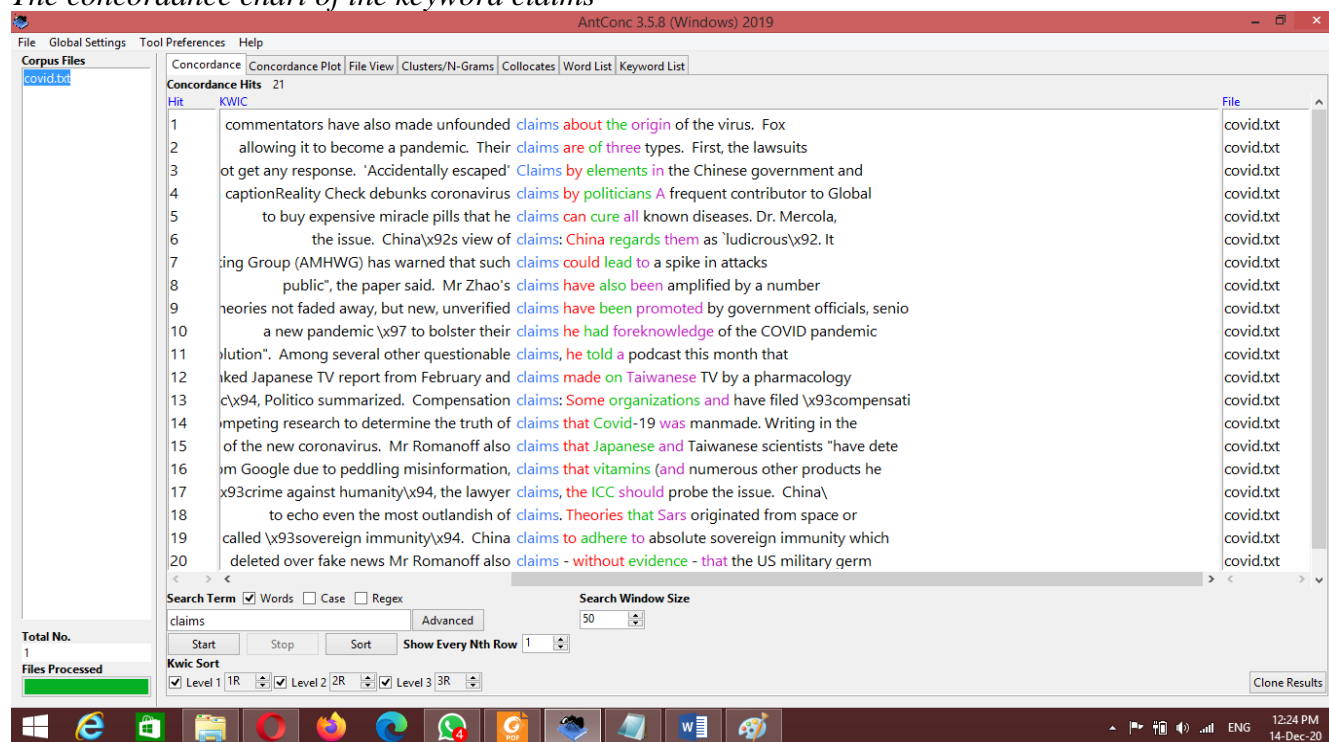
Collocates of the keyword claims

Rank	Frequency	Collocates
1	10	the
2	08	X
3	07	To
4	07	That
5	06	Of

Moving on, the word *claims* is the next most significant keyword investigated through N-Grams/Clusters (Table 9) and collocate list (Table 10) in the conspiracy theories corpus. Here, we can say that all forms of prepositions, determiners and articles are present in the examination of the word *claims* including the word *china*, and *pandemic*. In addition, claims by *politicians* and *elements* are also present in the concordance chart in Fig 2 shown below. So, the co-occurrence of all these words inform us about the debate going on the media regarding covid-19.

Figure 2

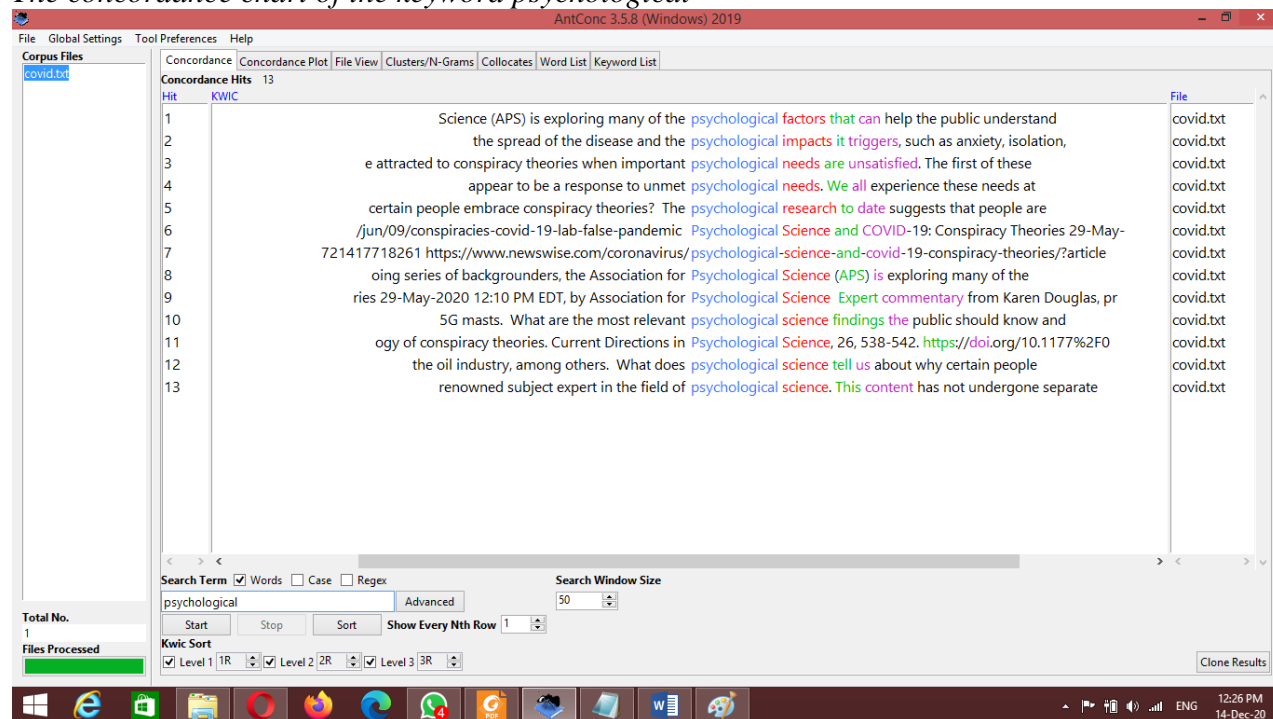
The concordance chart of the keyword claims



The next keyword analysed for concordance, N-Gram/Clusters and collocates is psychological (Tables 11 & 12) and it was found out that it occurred most frequently in concordance table as *psychological factors*, *psychological impacts*, *psychological needs* and *psychological research* (Fig. 3). In addition, the N-Grams results shows that all these words are exactly repeated in again and the collocates results is slight different in which *science*, *theories* and *conspiracy* is listed among top five words. The co-occurrence of psychological with factors, impacts, needs, research and science is confirmed in statistical analysis.

Figure 3

The concordance chart of the keyword psychological



The strongest predictors of beliefs in these ideas are a psychological predisposition to reject expert information and accounts of major events (denialism), a psychological predisposition to view major events as the product of conspiracy theories (conspiracy thinking), and partisan and ideological motivations Uscinski et al. (2020).

Table 11

Clusters/N-Grams of the keyword psychological

Rank	Frequency	Clusters
1	07	psychological science
2	02	Psychological needs
3	01	Psychological factor
4	01	Psychological impact
5	01	Psychological research

Table 12
Collocates of the keyword psychological

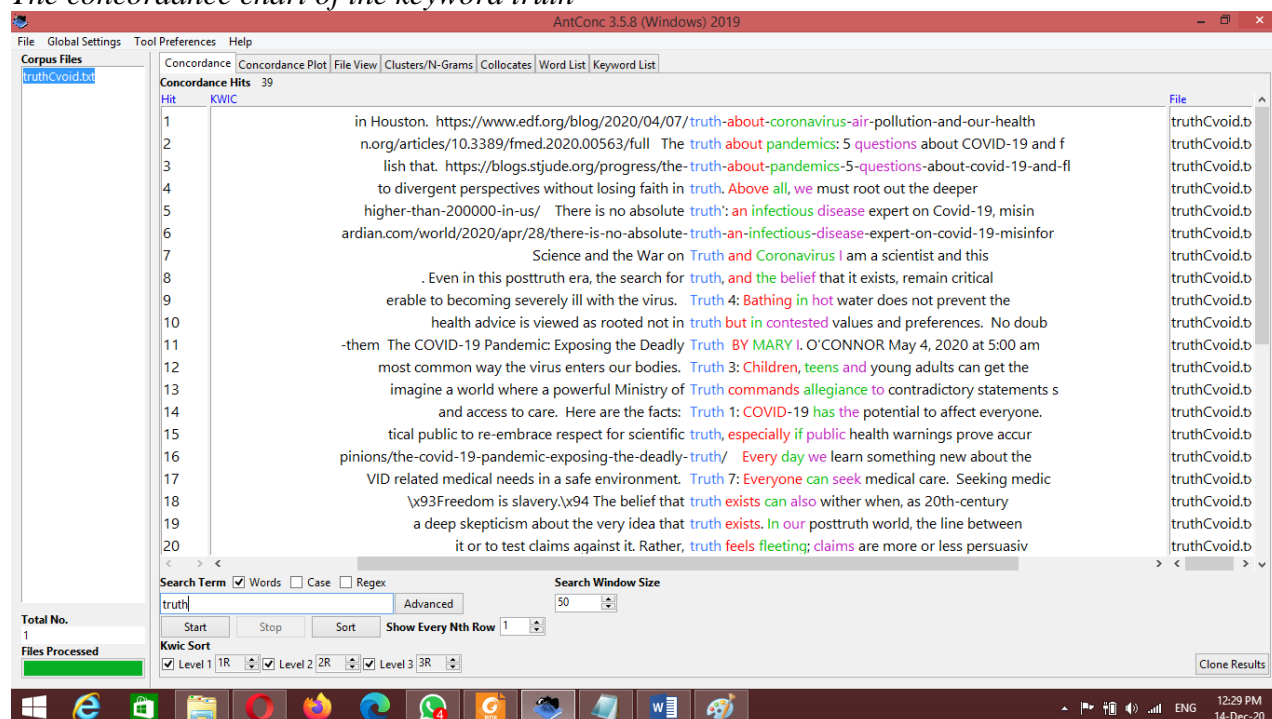
Rank	Frequency	Collocates
1	10	The
2	08	science
3	05	Theories
4	05	conspiracy
5	04	of

State-of-the-science methods should be deployed to debunk or displace the conspiracy theories and vaccine fears identified here (see Lewandowsky et al., 2012).

4.4 Truth based Covid

There are two important words in the keywords list (see Table 6) that need attention and further investigation one is *truth* and the second word is *post truth*. First, because *truth* is on the top of the list and second that *post truth* is an adjective and stands out from all the rest of the nouns, prepositions and conjunctions in the list. In the next section these two words will be processed further for collocation and concordance in order to see what sort of entailments these words have got in this context.

Figure 4
The concordance chart of the keyword truth



The next important keyword is truth for which concordance is applied and the list includes (Fig. 4 above), *truth about*, *truth above*, *truth an*, *truth and*, *truth but*, *truth by*, *truth especially*, *truth every day*, *truth exists*, and *truth feels*. The N-Grams/Clusters of the keyword truth (Table 13) include *truth in*, *truth exists*, *truth to*, *truth about* and *truth by*, and the collocates operation confirms (Table 14) *truth about* being used for 08 times in the list. It is followed by the word *theory* and *science*. So, it can be deducted that the co-occurrence of the keyword *truth* remains within the locus of these four words, i.e. *what the truth of pandemic is about*, moreover there is a correlation of truth, theory and science simultaneously occurred in the list.

Table 13

Cluster/ N-Grams of the keyword Truth

Rank	Frequency	Cluster
1	03	Truth in
2	02	Truth exists
3	02	Truth to
4	02	Truth about
5	02	Truth by

The N-Grams/Clusters of the keyword truth include *truth in*, *truth exists*, *truth to*, *truth about* and *truth by*. There is again a measure of relationship between the keyword truth with different prepositions, like in, about, by, about, and to.

Table 14

Collocates of the keyword Truth

Rank	Frequency	Collocates
1	21	The
2	13	X
3	12	That
4	09	To
5	09	Is
6	09	In
7	08	About
8	06	Some
9	05	Theory
10	05	Science

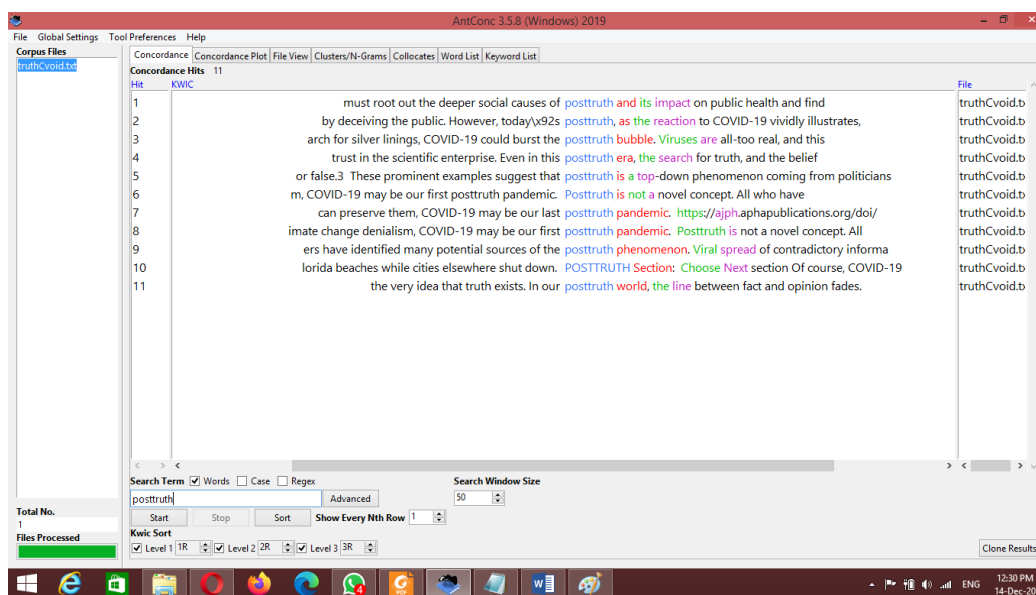
Another critical aspect of science is that a “truth” or “fact” in science is an evidence-based statement, not just a “subjective” feeling or an impression (Dobson, 2005). Wittgenstein believed that truth can be found through investigation of the everyday language games which may work in daily activities and all sort of social practices possible. Whereas, the famous thinker and

philosopher Nietzsche, wrote in his book *Will to Power* that truth is a form of error which a type of living species require in order to survive on everyday life. Therefore, truth is an inevitable part of our everyday lives and human beings do take its support in all situations and crisis including covid as suggested by the results of the concordance of the keyword *truth*. The collocational operation confirms the preposition *truth about* has some role in the corpus text. Moreover, the other two collocational words *theory* and *science* corresponds with the earlier results of the keywords.

The concordance of the post truth based corpus (Fig. 5 below) includes *post truth and*, *post truth as*, *post truth bubble*, *post truth era*, *post truth is*, *post truth pandemic*, *post truth phenomenon*, *post truth section*, *post truth world*. The N-Gram/Cluster result shows that truth exists occurred a couple of times whereas the rest of the words are prepositions (Table 15). The cluster/N-Gram of the keyword post Ruth include *post Ruth is*, *post truth pandemic*, *post truth and*, *post truth bubble*, and *post truth era*. Likewise, the collocates of the keyword *post truth* include *the*, *our*, *of*, *covid* and *pandemic* (Table 16). So, the standout words coming out from this analysis are, post truth bubble, post truth pandemic, and post truth era which includes in all three tables of the keyword. Last but not the least, it is statistically confirmed that whenever post truth word comes in the truth based corpus it is surrounded by era, bubble and pandemic.

Figure 5

The concordance chart of the keyword post truth



At a more general level, the market in expectation management is emblematic of the post-truth condition (Fuller 2020). The good news is that trust in science remains high, at least relative to other institutions. Moreover, if we search for silver linings, COVID-19 could burst the post truth bubble (Wendy 2020).

Table 15*The Cluster/N-Gram of the keyword posttruth*

Rank	Frequency	Cluster
1	02	Posttruth is
2	02	Posttruth pandemic
3	01	Posttruthand
4	01	Posttruth bubble
5	01	Posttruth era

It is an ideal medium for hype, exaggeration, falsehood, lies and gossip that are characteristic of the age of post-truth (Peters et al., 2018). Therefore, the clusters bubble, era and pandemic signals towards the same type of hype, exaggeration and falsehood hinted by Peters in his study. The cluster/N-Gram of the keyword post Ruth include *post Ruth is*, *post truth pandemic*, *post truth and*, *post truth bubble*, and *post truth era*.

Table 16*The Collocates of the keyword post truth*

Rank	Frequency	Collocates
1	06	The
2	04	Our
3	04	Of
4	04	Covid
5	03	pandemic

5. Conclusion

As the catastrophic round of the pandemic continues across the globe in the form of a nightmare for the world masses but the agencies responsible for the influx of information regarding covid-19 are divided into conspiracy theory and truth based reporting. This article had attempted to uncover all the salient features of both the sides by comparing the collected corpus and came out with the lexical differences and similarities which are being portrayed through these news stories. The results are interesting and beyond the level of mere intuition and can surprise the audience/readers. The study was only limited to the selected online newspaper stories during the first wave of covid 19 but in order to take more holistic view of the findings one needs to cover more span and should have a larger corpus.

The emergence of frequent words in conspiracy theory corpus hinted towards a trajectory build up linked with the newspapers and media regarding covid-19. The first and foremost result of the analysis confirmed the existence of a pandemic that is caused by a virus and related conspiracy theories surrounding with the pandemic. On the other hand, the truth-based corpus hinted towards the death of the masses around the globe during the pandemic that approximately mounts to 2.7 million and the counts goes on. The pandemic situation has also to do something

with the *science* that may have contributed or failed to contribute in the endemic. These words in the conspiracy corpus include *conspiracy*, *virus*, *theories* and *pandemic*, while the truth-based covid frequencies include *pandemic*, *death*, and *science*.

The unusually frequent occurring keywords *conspiracy*, *claims* and *psychological* from the conspiracy list and *truth* and *post truth* from truth based list. It is not surprising that *conspiracy* is the top rated word in conspiracy list and *truth* remains top rated keyword in the truth based list. The keyword *conspiracy* remained associated with *theory(s)* and *believer(s)* that suggested the fact conspiracy theory is successful owing to its believers. In addition, the keyword *psychological* suggested that the impact of the conspiracy theory had deep psychological underpinnings for the masses. The co-occurrence of the word *claims* is associated with *politicians* and certain *elements* who are perpetually disseminating the discourse of conspiracy therefore, it got rampant in the online stories because of harping the same theories time and again. In this context covid-19 is the conspiracy and it does not exist like other conspiracy theories of the present time including QAnon.

It is also paramount that scientists, health officials, world leaders and all those concerned must come up with the truth about covid 19 pandemic that can ultimately lead us towards a better understanding and empowering solution to end this pandemic. The building scenario suggests that the whole world is passing through severity of the situation and this may be turned into as the very first *post truth* pandemic phenomenon as suggested by the results and language/discourse has got a vital role to play in this informational jugglery.

References

- Anthony, L. (2019). AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>.
- Allport, G. W., & Postman, L. J. (1981). *The psychology of rumor*. New York: Henry Holt & Co.
- Baker, P. (2005). *Public Discourses of Gay Men*. London: Routledge.
- Bandari, R., Bandari, R, Zhou, Z, Qian, H, Tangherlini, TR, & Roychowdhury, VP. (2017). A *Resistant Strain: Revealing the Online Grassroots Rise of the Antivaccination Movement*. (COMPUTER, 60-67.) eScholarship, University of California.
- Cheng, W. (2009). Income/interest/net: Using internal criteria to determine the aboutness of a text In K. Aijmer (Ed.), *Corpora and Language Teaching* (pp. 157-178). Amsterdam, the Netherlands: John Benjamins.
- De, C. M., Giard, L., & Mayol, P. (2014). *Practice of Everyday Life: Volume 2: Living and Cooking*. Minneapolis: University of Minnesota Press.
- Dobson GP. (2005). *A Chaos of Delight: Science, Religion and Myth and the Shaping of Western Thought*. Oakville, CT: David Brown and Co.
- Fine, G. A., & Turner, P. A. (2004). *Whispers on the color line: Rumor and race in America*. Berkeley, Calif: University of California Press.
- Fuller, S. (2020). *When a virus goes viral: Life with COVID-19*. Social Epistemology Review and Reply Collective, 17 March. <https://social-epistemology.com/2020/03/17/when-a-virus-goes-viral-life-with-covid-19-steve-fuller/>. Accessed 13 April 2020.
- Kitta, A. (2012). *Vaccinations and public concern in history: Legend, rumor, and risk perception*. New York, NY: Routledge.
- Lewandowsky et al., (2012) *Misinformation and its correction: continued influence and successful debiasing*. Psychol. Sci. Publ. Interest, 13 (2012), pp. 106-131, 10.1177/1529100612451018
- Mahlberg, M., & McIntyre, D. (2011). A case for corpus stylistics: Ian Fleming's *Casino Royale*. *English Text Construction*, 4(2), 204-227.
- National Commission on Terrorist Attacks upon the United States., Kean, T. H., & Hamilton, L. H. (2004). *The 9/11 Commission report: Final report of the National Commission on Terrorist Attacks upon the United States*. New York: Norton.

- In Campion-Vincent, V., In Fine, G. A., & In Heath, C. (2017). *Rumor Mills: The Social Impact of Rumor and Legend*.
- Hoey, M. (1986) 'The discourse colony: A preliminary study of a neglected discourse type': 1–26 in *Talking about Text*, Discourse Analysis Monograph no. 13, English Language Research, University of Birmingham.
- Peters, M. A., Rider, S., Hyvoenen, M., & Besley, T. (Eds.). (2018). *Post-truth, fake news: Viral modernity & higher education*. Springer.
- Rosnow, R. L. (January 01, 1980). Psychology of rumor reconsidered. *Psychological Bulletin*, 87, 3, 578-591.
- Scott, M. and Tribble, C. (2006). *Textual Patterns: Key words and corpus analysis in language education*. Amsterdam: John Benjamins.
- Shibutani, T. (1977). *Improvised news: A sociological study of rumor*. Indianapolis: Bobbs-Merrill educational Publishing.
- Shalom, C. (1997) 'That great supermarket of desire: Attributes of the desired other in personal advertisements': 186–203 in Harvey, K., Shalom, C. (eds) *Language and Desire*. London: Routledge.
- Sinclair, J. M. (2005). *Document Relativity*. Tuscany, Italy: Tuscan Word Centre.
- Slotkin, J. (2020). U.K. Cellphone Towers Ablaze As Conspiracy Theories Link 5G Networks To COVID-19. NPR.org. Retrieved from [www.npr.org website:
https://www.npr.org/sections/coronavirus-live-updates/2020/04/04/827343675/u-k-cellphone-towers-ablaze-as-conspiracy-theories-link-5g-networks-to-covid-19](https://www.npr.org/sections/coronavirus-live-updates/2020/04/04/827343675/u-k-cellphone-towers-ablaze-as-conspiracy-theories-link-5g-networks-to-covid-19)
- Swire, B., Berinsky, A. J., Lewandowsky, S., & Ecker, U. K. H. (2017). Processing political misinformation: comprehending the Trump phenomenon. *Royal Society Open Science*, 4(3). doi:10.1098/rsos.160802
- Tangherlini, T. R. (2018). Toward a generative model of legend: Pizzas, bridges, vaccines, and witches. *Humanities*, 7(1), 1
- Uscinski, J. E. (2019). *Conspiracy theories and the people who believe them*.
- Uscinski, J. E., Enders, A. M., Klostad, C., Seelig, M., Funchion, J., Everett, C., Wuchty, S. Murthi, M. (April 28, 2020). Why do people believe COVID-19 conspiracy theories?. *Harvard Kennedy School Misinformation Review*.
- Victor, J. S. (1993). *Satanic panic: The creation of a contemporary legend*. Chicago: Open Court.

Watson Todd, R. (2013). Identifying new knowledge in texts through corpus analysis.
International Journal of Language Studies, 7(4), 57-76.

Wendy E. Parmet, Jeremy Paul, (July 1, 2020) *COVID-19: The First Posttruth Pandemic*,
American Journal of Public Health 110, no. 7: pp. 945-946.

Appendix

References of Online [selected] Newspaper Stories

Javed. (2020, June 10). *China rejects conspiracy theories about origin of COVID-19*. INP
<https://www.inp.net.pk/china-rejects-conspiracy-theories-about-origin-of-covid-19/>

Sardarizadeh, S. and Robinson, O. (2020, April 26). *Coronavirus: US and China trade conspiracy theories*. BBC
<https://www.bbc.com/news/world-52224331>

Javed, M. (2020, April 29). COVID19 and conspiracy theories: A Chinese virus or a bio-weapon? Modern diplomacy.
<https://modern diplomacy.eu/2020/04/29/covid19-and-conspiracy-theories-a-chinese-virus-or-a-bio-weapon/>

Kuo, A. M. (2020, May 19). *Investigating China: COVID-19 and the CCP*. The Diplomat
<https://thediplomat.com/2020/05/investigating-china-covid-19-and-the-ccp/>

Aydin, S. (2020, June 19). Britain's contemporary Islamophobia spawns COVID-19 conspiracy theory. Daily Sabah.

<https://www.dailysabah.com/opinion/op-ed/britains-contemporary-islamophobia-spawns-covid-19-conspiracy-theory>

Lynas, M. (2020, April 20). COVID: Top 10 current conspiracy theories. Alliance for Science.
<https://allianceforscience.cornell.edu/blog/2020/04/covid-top-10-current-conspiracy-theories/>

Dasak, P. (2020, June 9). Ignore the conspiracy theories: scientists know Covid-19 wasn't created in a lab. The Guardian.

<https://www.theguardian.com/commentisfree/2020/jun/09/conspiracies-covid-19-lab-false-pandemic>