

Ewa Kusz (Rzeszów)\*  
ORCID: 0000-0002-5245-419X

Received: 15.12.2020  
Accepted: 6.05.2021  
Published: 23.12.2021

## Disinformation, misinformation and fake news in the time of the Covid-19 pandemic. A corpus-based approach

The Covid-19 outbreak and its dissemination resulted in the creation of a new, specialised and global discourse among the population. This paper aims to investigate a tsunami of dis- and misinformation on the basis of selected, most frequently occurring, items of fake news to show how the ‘infodemic’ (Covid-19-related misinformation) has expanded, and how fake news is structured. The paper concludes by presenting patterns of fake news, including grammar structures, the frequency of the given lemmas and collocations, in the hope that it will provide greater transparency, help to flatten the ‘infodemic’ curve and make the readers more aware of how myths about coronavirus have been formed during the pandemic.

**Keywords:** fake news, misinformation, disinformation, Covid-19, coronavirus, corpus-based approach, linguistic features

### Desinformation, Fehlinformation und Fake News in der Zeit der Covid-19-Pandemie. Ein korpusbasierter Ansatz

Der Ausbruch von Covid-19 und seine Verbreitung haben einen neuen, spezialisierten und globalen Diskurs in der Bevölkerung ausgelöst. In diesem Beitrag soll ein Tsunami von Des- und Fehlinformationen anhand ausgewählter, am häufigsten vorkommender Fake News untersucht werden, um zu zeigen, wie sich die „Infodemie“ (Covid-19-bezogene Fehlinformationen) ausgeweitet hat und wie Fake News strukturiert sind. Abschließend werden Muster von Fake News präsentiert, einschließlich grammatischer Strukturen und der Häufigkeit des Auftretens von ausgewählten Lemmata und Kollokationen. Dabei sollen die Möglichkeiten aufgezeigt werden, wie für mehr Transparenz gesorgt und wie die „infodemische“ Kurve abgeflacht werden kann. Den Lesern soll dabei

---

\* dr Ewa Kusz, Uniwersytet Rzeszowski, Katedra Lingwistyki Stosowanej, al. mjr. W. Kopi-  
sto 2 B, 35- 315 Rzeszów, e-mail: e.lamasz@gmail.com

bewusst gemacht werden, wie sich während der Pandemie Mythen über das Coronavirus gebildet haben.

**Schlüsselwörter:** Fake News, Fehlinformation, Desinformation, Covid-19, Coronavirus, corpus-basierter Ansatz, linguistische Merkmale

## Dezinformacja i fake news w dobie pandemii Covid-19. Podejście korpusowe

Wybuch epidemii Covid-19 i jej rozprzestrzenianie się spowodowały powstanie nowego, specjalistycznego i globalnego dyskursu wśród społeczeństwa. Celem artykułu jest zbadanie swoistego tsunami dezinformacji na podstawie wybranych, najczęściej pojawiających się fake newsów, aby pokazać jak rozszerzyła się „infodemia” (dezinformacja związana z Covid-19) i jakie struktury pojawiają się przy tworzeniu tego typu komunikatów. Artykuł kończy się prezentacją wzorców fake news’ów, w tym struktur gramatycznych, częstotliwości występowania danych lematów i kolokacji, w nadziei, że zapewni to większą przejrzystość, pomoże spłaszczyć krzywą „infodemii” i uświadomi czytelnikom, jak w czasie pandemii powstawały mity na temat koronawirusa.

**Słowa kluczowe:** fake news, dezinformacja, mylna informacja, Covid-19, koronawirus, podejście korpusowe, cechy lingwistyczne

## 1. Introduction

The word *coronavirus* may well take the crown as the world’s most well-known internationalism of 2020, having gained its global position within just a few months after March 2020, when the World Health Organisation (WHO) announced the COVID-19 pandemic. In fact, the emergence of a new internationalism is only the tip of the iceberg that begins a global discourse about the pandemic and massive miscommunication related to SARS-CoV-2. In response to that, the director-general of the WHO in a speech in February 2020 highlighted the importance of combating the spread of disinformation, misinformation and fake news, by saying that “[it] ... might be the most contagious thing about [coronavirus]” and “WHO’s ‘infodemics’ team is working hand in glove with our communication department to deliver information to a broader public audience”<sup>1</sup>. Indeed, the proliferation of fake news about COVID-19 is unprecedented and requires some indispensable arrangements in order to stop the spread of misleading information, false beliefs and misconceptions, which may be harmful to society by fuelling panic and threatening public health. Recent studies have shown that there are several factors that may contribute to a person’s susceptibility to

<sup>1</sup> Director-General’s remarks at the media briefing on the 2019 novel coronavirus on 8 February 2020. <https://www.who.int/director-general/speeches/detail/director-general-s-remarks-at-the-media-briefing-on-2019-novel-coronavirus---8-february-2020>. Accessed November 2, 2020.

misinformation, including low level or lack of education (Piller et al. 2020). Other detrimental effects observed in the time of the infodemic were shown in a Snopes' analysis which reported that one in three people in such countries as Argentina, Germany, South Korea, Spain, the UK and the USA have seen myths and fake news about COVID-19 on social media and links/videos shared on such platforms as Facebook, Google, YouTube, Twitter, etc. (Porter 2020). In fact, there is a large number of people whose only source of news is the links posted on social media (Allcott and Gentzkow 2017, Thompson et al. 2019 and Ku et al., 2019). As a result, the abovementioned platforms, as well as many others, have started to collaborate with the WHO by developing policies to eliminate disinformation and misinformation about coronavirus. Though the existence of fake news is not novel, it is now alarming due to the speed at which those false beliefs are perpetuated.

To respond to this call, our research study aims to observe the fight against 'corona-myths' and attempts to arm the readers with models of fake news in the hope that it enables them to differentiate misleading information from valuable sources. Also, following the recent research on the proliferation of fake news about COVID-19, we aim to find out why many people believe that such things as drinking bleach or rinsing with salty water could prevent or cure coronavirus (Lamos et al. 2020). The study focuses on the analysis of one thousand of the most popular items of fake news, which have been debunked on the WHO website and other frequently visited authoritative websites concerning news about COVID-19. By using Sketch Engine®, a special corpus programme, we carefully analysed the linguistic mechanisms underlying 'corona-based' fake news, including content words, language structures and recurring verbs, and we distinguished several constructs to enable exploration of which categories were most frequently chosen. One thousand items of fake news were subsequently juxtaposed and compared with the same number of factual stories about coronavirus. The novelty of this research is that it extends the literature on some linguistic features in terms of misconceptions about coronavirus during the COVID-19 pandemic.

The study is structured as follows. In the literature review section, we attempt to present works that focus on the sharing of dis- and misinformation, as well as the phenomenon of fake news. Prior to forming the research questions and hypotheses, we focus on the theoretical background regarding the sharing of fake news. The next part describes the methods and results, whereas the analysis and interpretation of results are presented in the discussion section. Limitations and future work constitute the final part of this work.

## 2. Literature review

The existence of fabricated information is not new; however, as Nougayrède (2018) highlights, never before has there been technology so advanced and easily-accessible as the Internet is today. This results in fast dissemination of news, both true or fake. Knowing the historical context of propaganda may be crucial to understand the reason of today's 'infodemic' and what Wardle and Derakhshan (2017) call the 'information disorder'.

Seeking for the first signs of mis- and disinformation, we can go back to Roman times, when Octavian spread false information about Antony to ruin his reputation and to become the first Roman Emperor. As Kaminska (2017) notes, "fake news allowed Octavian to hack the republican system once and for all".

The invention of the printing press in the 1440s contributed to the proliferation of dis- and misinformation. A growing number of conflicts in the following centuries, including wars, catastrophes and changes in the systems of government became even greater markers for the dissemination of propaganda. It was the advent of radio in the late 19<sup>th</sup> century and television in the first half of the 20<sup>th</sup> century that led to the evolution of communication technology. Finally, the arrival of the Internet in the late 20<sup>th</sup> century and social media in the 21<sup>st</sup> century contributed to fake news dissemination. As Posetti and Matthews (2018: 1) emphasise:

[w]e now inhabit a world with computational propaganda, state-sponsored 'sock-puppet networks', troll armies, and technology that can mimic legitimate news websites and seamlessly manipulate audio and video to create synthetic representations of any number of sources. In this environment, where trust becomes polarised around what "news" aligns with their views, many news consumers feel entitled to choose or create their own 'facts'.

As a result, we are not surprised to hear that information is a 'weapon' prepared by politicians, governments, secret services or counterintelligence agencies to inflame international tensions or to start hostilities. Moreover, the dissemination of fake news is a 'weapon' used to attack and eliminate other political competitors (e.g. fake news referring to Hillary Clinton). Also, there are items of news that have a significant impact on international relations (e.g. fabricated news about Pope Francis' support for Donald Trump's candidacy during the US election in 2016) (Antos 2019; Palczewski 2019).

As seen above, understanding the phenomenon of fake news and being able to distinguish it from real news can be an antidote to the proliferation of fabricated information.

## **Fake news**

Although there are many definitions that provide an explanation of the concept of fake news, there is one key phrase that can be found in most of them, i.e., that it is misleading information. For instance, McGonagle (2017) emphasises that fake news is deliberately fabricated information whose main aim is to misinform and deceive the receivers. Similarly, Duffy et al. (2019) define fake news as any information that includes misleading content and mimics reliable facts. A more complex explanation of fake news is provided by Vasu et al. (2018: 5) who state that it should be understood as ‘a medium for a spectrum of phenomena comprising the following categories’, i.e., disinformation, misinformation (two types, see below), entertainment and ‘falsehoods distributed for financial gain’ (2018: 5). Following that, Apuke and Omar (2020a) and Wang et al. (2019) emphasise that fake news is untrue information, spread deliberately or unintentionally on social media. This includes myths, rumours, conspiracy theories and/or hoaxes. While the process of sharing fake news can be unintentional, as people may not be aware of the fact that they are posting untrue information, its creation is deliberate (Egelhofer and Lecheler 2019). In this section, we cover three concepts, i.e., disinformation, misinformation and fake news in general, and we attempt to explain their phenomena on the basis of the recent studies on the dissemination of fake news regarding COVID-19.

## **Misinformation and disinformation**

Just as with fake news, the term misinformation has a large number of definitions. Vasu et al. (2018: 5) provide two ways to understanding this phenomenon. The first explains that it is ‘falsehoods and rumours propagated as part of a political agenda...based on ideological bias’, whereas the second definition states that it is intentionally or unintentionally proliferated without a certain political aim, yet it achieves viral status. Scheufele and Krause (2019) define misinformation as incorrect information, which most probably is accidentally produced. Similarly, Wardle and Derakhshan (2017) point out that misinformation occurs when a false belief is being shared unintentionally and without the intention to cause harmful effects. Thus, misinformation is usually disseminated by well-meaning but ill-informed people (Howell 2020).

On the contrary, according to Oxford Languages, disinformation is understood as ‘false information, which is intended to mislead, especially propaganda issued by a government organisation to a rival power or the media’. While both terms, i.e., mis- and disinformation refer to forms of inaccurate information (Bastick 2020), according to Pal and Banerjee (2019), it is disinformation that is aimed at deceiving the public.

### **Studies of disinformation, misinformation and fake news related to COVID-19**

According to Apuke and Omar (2020), there is a growing, though still limited, number of studies that focus on fake news. Recent research has shown that items of fake news shared within the last few months of 2020 (excluding December 2020) delivered untrue information mainly about COVID-19 and mostly, due to lockdown, they were shared online (Pennycook et al. 2020). The CoronaVirus Fact Alliance database recorded approximately 4,000 COVID-19-related fabricated news items in April 2020 (Howell 2020). Waszak et al. (2018) highlight the fact that the popularity of social media encourages people to share information (either reliable or unreliable) without restriction and thus proliferates false content, which has a detrimental effect, especially on health services, stirring panic and undermining medical advice (Apuke and Omar 2020, Pulido et al. 2020). Apuke and Omar (2020) in their study modelled the predictors of fake news sharing among social media users. Their results revealed that there is a correlation between the concept of altruism, which appeared to be the most significant factor, and the prediction of sharing fake news about COVID-19. In Bastick's (2020) pre-proofed paper, he found that fake news can covertly modify a reader's behaviour. He emphasises that the current methods of fighting against the proliferation of fake news are insufficient and there is a strong need to make an inter-disciplinary effort to protect especially social media users who are mainly exposed to mis- and disinformation. In line with that, Islam et al. (2020) developed a model in which they investigated motifs that among others influence the sharing of fake news as well as social media fatigue. Barua et al. (2020) also admit that the dissemination of fabricated information about COVID-19 on social media platforms is now faster than the spread of the virus itself. Hence, in their study, they considered three types of misconceptions as the stimulus, i.e., general misinformation, conspiracy and religious misinformation beliefs, suggesting credibility evaluations as a resilience strategy. Finally, they tested the obtained effects on individual responses to COVID-19-related fake news. Their findings revealed that credibility evaluation of fake news is a significant predictor of individual responses to false coronavirus stories (Barua et al. 2020: 7).

The dissemination of false beliefs on social media platforms and other online sources is not the only issue taken into consideration in the studies regarding the phenomenon of COVID-19-related fake news. Piller et al. (2020) describe the language challenges concerning a multilingual crisis of communication. They emphasise the dominance of English-centric global mass communication and thus the devaluation of 'minoritised languages' (Piller et al. 2020: 503). On the one hand, their study aims to explore the sociolinguistics of the challenges presented by COVID-19 and on the other to 'open a space for intercultural dialogue within sociolinguistics' (Piller et al. 2020: 503). There is one more study that is worth

mentioning in terms of COVID-19-related fake news and the linguistic features and methods employed. In their study, Tan et al. (2020) explain that most of the research conducted on the proliferation of fake news in the time of the COVID-19 pandemic focus on the psychological and other deleterious effects rather than language challenges or linguistic trends. Considering that the tools and procedures of news verification are still slow and time-consuming, there is a strong need to provide the public with a reliable self-verification tool to be used in emergency situations (Tan et al. 2020: 2). Hence, they examined the concept of ‘fear’ and its linguistic realisations on the basis of COVID-19-related fake news and compared them with reliable information in order to investigate the linguistic properties of fake news. Brennen et al. (2020) identify some of the main types, sources and claims of COVID-19 misinformation by analysing fake news published in English between January and the end of March 2020. Their analysis revealed that there is a vast number of various forms, sources and claims included in COVID-19 misconceptions, which are mostly reconfigured rather than completely fabricated.

In line with the previous studies and responding to the gaps in the literature, the current research project presents predictive models that include linguistic factors that can help in the detection and verification of fake news.

### **Research questions and hypotheses**

The study was guided by two research questions:

RQ1: Which parts of speech are the most frequently used in the given sample of fake news?

RQ2: Considering corpus-based analysis, is it possible to create a model of fake news?

Based on the previous studies concerning fake news in the time of the COVID-19 pandemic, and the research questions, the following hypotheses are presented:

H1: There is potential for a model of COVID-19 -related fake news which can be built on the basis of the analysis of the given examples

H2: The created models enable the readers to differentiate fake news from real information

### **3. Methodology**

This section describes the method used to conduct the research. The model of this study was developed with a corpus-based methodology on the basis of which the following steps were taken: corpus development, detailed analysis and



interpretation of the obtained results. One of the main strengths of the corpus-based approach is that it provides a database of naturally-occurring discourse. This enables empirical analysis of the actual patterns that exist in a particular language. Secondly, a corpus-based methodology, together with (semi-) automatic computational tools, gives detailed and comprehensive data.

### **Data collection**

The corpus used in this study includes one thousand items of fake news the credibility of which was denied by the WHO on their website as well as COVID-19: Poynter Resources, and several frequently visited reliable online sources which aimed to demonstrate and explain the inaccuracy of the most common misconceptions about coronavirus. The database gathered items of fake news that appeared in more than 70 countries and about 40 languages. One thousand items of fake news were also compared to the same number of factual stories that are available online. Our rationale for using such a corpus is that the given false beliefs are those which seem to be the most popular among the public. A detailed analysis of their linguistic features, including content words, structures used and recurring verbs, enable us to build a model of fake news. Also, prior to the interpretation of the obtained results, we distinguished several constructs in order to explore the most frequently used categories that were established regarding their key words and the main topics in which fake news occurs. Twenty-one categories were distinguished:

- age (e.g. ‘Coronavirus only affects older people’);
- animals (e.g. ‘Video footage of dead animals such as dogs, rats, bats and snakes being sold in a market in Wuhan’; ‘The COVID-19 virus can be spread through mosquito bites’);
- clothes (e.g. ‘COVID-19 spreads through shoes and clothes’);
- conspiracy theory (e.g. Bill and Melinda Gates Foundation & others predicted up to 65 million deaths via coronavirus – in simulation run 3 months ago’; ‘There is a correlation between 5G technology and the spread of Sars-CoV-2’);
- cure (e.g. ‘Vitamin and mineral supplements can cure COVID-19’);
- danger (e.g. ‘People living with HIV are more likely to get seriously ill’);
- death (e.g. ‘Doctors said the entire Wuhan population will die of the novel coronavirus’);
- detection (e.g. Digital thermometers are 100% effective in detecting COVID-19 patients’);
- distance (e.g. ‘Coronavirus has a reach of up to 8 meters’);
- food (e.g. ‘A seafood market is the source of the novel coronavirus, strange animals and foods were sold there, such as bat soup and snakes’);



- origin (e.g. ‘The outbreak was a result of an accidental release from a Wuhan laboratory’);
- political actions (e.g. ‘Sars-CoV-2 was created by governments’; ‘The Chinese government had created the virus’; ‘The US government created the virus to undermine the Chinese government’);
- precaution (e.g. ‘Avoid ice-creams, cold drinks and sweets for 90 days to prevent coronavirus’),
- race (e.g. ‘People of colour may be immune to the coronavirus because of melanin’);
- side-effects (e.g. COVID-19 is increasing the risk of stroke in infected patients’);
- testing (e.g. ‘Holding your breath for ten seconds is a test for SARS-CoV-2’);
- time (e.g. ‘You have to be with someone for 10 minutes to catch the virus’);
- vaccine (e.g. ‘People who have gotten sick with COVID-19 will not benefit from getting vaccinated’);
- virus transmission (e.g. ‘a package from AliExpress sent from China can spread the new coronavirus’),
- wearing a mask (e.g. ‘You have to wear a disposable mask with the blue side out if you are sick and the white side out if you are not’);
- weather (e.g. ‘the novel coronavirus [...] will not last long in the Philippines because of its warm climate’).

Additionally, we juxtaposed the given items of fake news with the same number of factual stories about coronavirus and attempted to compare their structures, as well as the frequency of categories to which they were assigned.

Sketch Engine® was chosen as the corpus program for this research project. It is a corpus tool that provides its users with the ability to study and analyse concordances, collocation, key words, word lists and word sketches using researchers’ own corpus. On the basis of that, one thousand items of fake news and one thousand facts were uploaded into Sketch Engine® which resulted in 34,470 words being obtained. It should be noted that the following analysis assumed that one sentence equals one fake news claim and one fact.

### **Data analysis and measures**

First, it is important to note that data analysis of the given list of fake news items and facts was a two-step process. In step 1, we distinguished several constructs thanks to which all of the misconceptions were grouped into certain categories. Subsequently, in step 2, using Sketch Engine®, we checked the frequencies of nouns, verbs, content words and structures used to create the items of fake news and facts. On the basis of that, it was possible to create certain models of the most frequently appearing fake news claims on the Internet, including social media, as well as to design the pattern for real information.

### 4. Results

For the results of step one, twenty-one categories were distinguished (Figures 1 and 2) among which two constructs, i.e. conspiracy theories about COVID-19 and the virus transmission, turned out to be featured in the most frequent fake news claims, constituting 18% and 16% of all examples, respectively. Across the same sample, the next most common constructs within pieces of fake news concern political actions (11%), death from COVID-19 (9%) and danger of COVID-19 (8%). The relatively high position of fake news regarding political actions may be due to public distrust in governments. Comparing these results with other data available, it is possible to notice a substantial discrepancy between outcomes obtained at the beginning of 2020 and in January 2021 (the current research). For instance, Brennen et al. (2020) reveal that COVID-19-related fake news that appeared from January to March 2020 concerned mainly actions of public authorities (39%) and the spread of COVID-19 through communities (24%). However, fake news about vaccine development constituted only 6% of other available constructs.

Our sample also includes three more constructs that should be emphasised due to their frequency in the proliferation of fake news, namely cure-related topics (5%), origin (5%) and fabricated information about precautions that must be taken (5%).

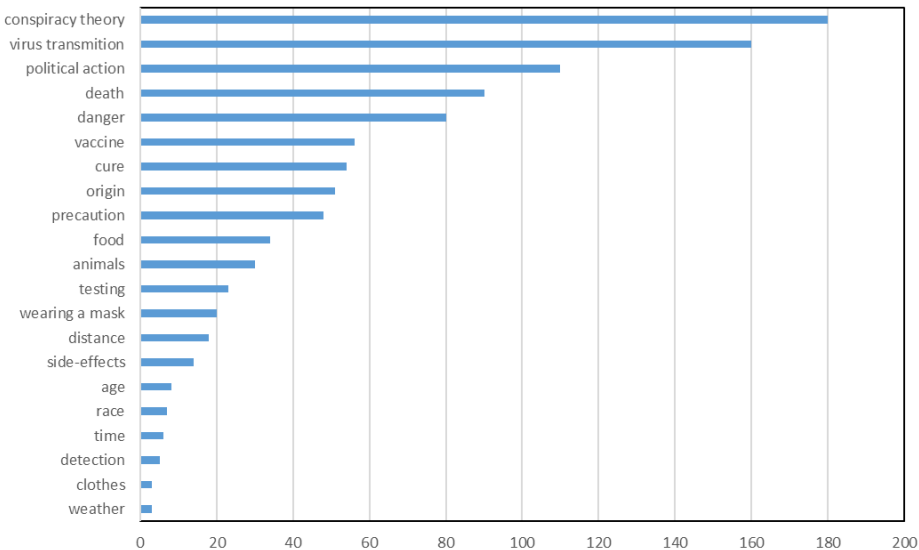


Figure 1. The distribution of fake news constructs.

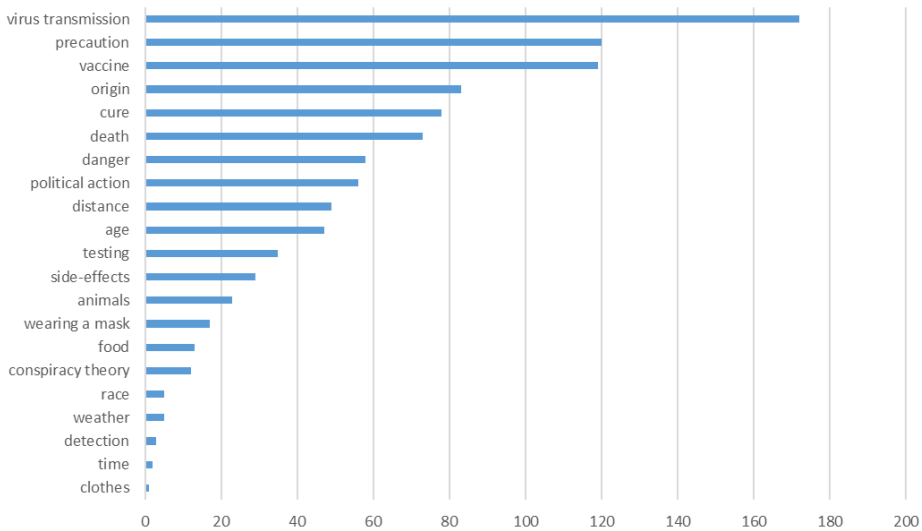


Figure 2. The distribution of real news constructs.

By comparing the distribution of fake news constructs with the real news, it can be easily noticed that virus transmission was the only category that appeared at the top of both lists. It took the first place among facts (17%) and second for fake news. This should not come as a surprise, as while Covid-19 spread rapidly around the globe, people strove to get more information about coronavirus and its transmission. Only two categories, namely detection (3% of facts, 1% of fake news) and race (5% of facts, 1% of fake news) were placed in the same positions; however, they were among the last constructs that appeared on the given lists.

For step 2, checking of the frequencies of content words, nouns, verbs and grammatical structures led to the following results being obtained. As shown below in Table 1, the most frequent word among the one thousand items of fake news investigated is the definite article 'the' with 1190 uses, followed by the noun 'coronavirus' (800), the indefinite article 'a' (605), and the preposition 'in' (730). The word 'new' and 'Wuhan' appeared 200 times each, whereas 'China' and 'people' 175 times. The word 'virus' (160) was placed in 13th position. 'Chinese' appeared 120 times, and words such as 'novel' 'infected' and 'outbreak' 115 times each. The analysis of concordance of the definite article reveals that 'the' collocates mainly with the noun 'coronavirus' (170), 'new' (145), 'novel' (70) and 'virus' (65).

<b>Word</b>	<b>Frequency</b>
the	1190
coronavirus	800
a	605
in	730
new	200
Wuhan	200

Table 1. List of the most frequently used words in the sample (fake news).

In order to create a pattern structure for the given fake news claims, we shall start with an analysis of the frequency of given words, starting from nouns (Table 2). Among all of the nouns that appeared in the given sample, ‘coronavirus’ is the most common (800 times), ‘Wuhan’ takes the second place (200) and ‘China’ is the next most frequent noun (175). The nouns ‘people’ (175), ‘virus’ (160), and ‘outbreak’ (115) come further down the list.

<b>Noun</b>	<b>Frequency</b>
coronavirus	800
Wuhan	200
China	175
people	175
virus	160

Table 2. The frequency list of nouns in the sample (fake news).

Another category worthy of investigating is the frequency of verbs that occur in the analysed sample (Table 3). Apart from the most common verb ‘to be’, which was used 630 times, ‘have’ was mentioned 160 times, ‘die’ 80 times and ‘spread’ 60, whereas ‘prevent’ took the sixth position. The other common verbs were ‘kill’ (59), ‘infect’ (57) and ‘create’ (47). As for adjectives, the words ‘new’ (200), ‘novel’ (115) and ‘infected’ (115) were the most popular, whereas for adverbs the word ‘not’ took first position. The most common pronouns were ‘it’ (85) and ‘its’ (30). There were 330 uses of the conjunction ‘and’, 45 of ‘or’. The most common prepositions were ‘in’ (730), ‘of’ (445) and ‘to’ (190).

<b>Verb</b>	<b>Frequency</b>
be	630
have	160

Verb	Frequency
die	80
spread	60

Table 3. The frequency list of verbs in the sample (fake news).

After examining the concordances, it can be noticed that the noun ‘coronavirus’ strongly collocates with the adjective ‘new’ (170) and ‘novel’ (115), as well as the definite article ‘the’ (405). Accordingly, the second most common noun ‘Wuhan’ mostly collocates with ‘in’ (120), ‘market’ (35) and ‘China’ (35). The noun ‘people’ goes with ‘died’ (30), ‘infected’ (20) and ‘China’ (30). Such a detailed examination of collocations among the most frequently used nouns in the given sample indicate a strong interdependence between the virus-oriented nouns and the place where it all started, namely Wuhan in China. This goes in line with the fact that the most common constructs among the analysed items of fake news focused on taking every precaution to prevent infection and information on what the virus could supposedly cause. However, this information does not seem very different than the real news use of language. It would not be possible to create a pattern structure for the given fake news claims if it was not compared to the forms used in the items of real news.

Word	Frequency
the	1123
coronavirus	914
a	589
COVID-19	446
China	118
Wuhan	110

Table 4. List of the most frequently used words in the sample (real news).

The analysis of the frequency of given words in terms of real news does not reveal many differences with the nouns that appeared in fake news (Tables 1 and 4). As shown in Table 4, the definite article ‘the’ with 1123 uses is the most common word among facts. ‘Coronavirus’ takes the second position with 914 uses, whereas the indefinite article ‘a’ (used 589 times) is in third place. Later, there are such nouns as ‘COVID-19’ (446), China (118), and ‘Wuhan’ (110). Following that, there is almost no difference between the structure patterns used to build fake news and facts. In both patterns, ‘the’ is the most common word, and ‘a’ is in third place. What we can distinguish from the comparative analysis of

both samples is the lack of adjectives in the pattern structures for facts, whereas the adjective ‘new’ is in the top five for the items of fake news. The main function of adjectives is to modify nouns, which might be the reason why they are more frequently used in the items of fake news than in facts.

Noun	Frequency
coronavirus	914
COVID-19	446
China	118
Wuhan	110
virus	87

Table 5. The frequency list of nouns in the sample (real news).

Among all of the nouns that occurred in the given sample of real news, ‘coronavirus’ is the most common (914 times), ‘COVID-19’ takes the second place (446) and ‘China’ is the next most frequent noun (118). The nouns ‘Wuhan’ (170), and ‘virus’ (87) come further down the list. The given samples presented in Tables 2 and 4 do not show significant differences between structure patterns for the items of fake and real news. However, the comparative analysis of the most frequently used verbs revealed that, besides the appearance of such common verbs as ‘be’ and ‘have’, the verbs ‘spread’ and ‘protect’ are in the leading positions among facts. On the contrary, the verb ‘protect’ does not appear in the sample referring to fake news, whereas ‘to die’ occurs 80 times and is in third place (see Table 3).

Verb	Frequency
be	616
have	207
spread	94
protect	52

Table 6. The frequency list of verbs in the sample (real news).

## 5. Discussion

The corpus-based approach implemented in the study enabled us to model the constructs that dominate in the corpus of 1000 items of fake news and 1000 items of real news gathered in November 2020. Specifically, we focused on the

dominance of certain claims. Conspiracy theories and the transmission of the virus were found to be the strongest predictors of the main constructs of fake news related to COVID-19, whereas virus transmission and taking precautions of real news. The picture which emerges from the given samples shows that most common conceptions provided by facts contain precautionary measures for curbing COVID-19. On the contrary, fake news focuses on conspiracy theories, virus transmission and political actions. Interestingly, reconfigured content to some extent contained true information (e.g. COVID-19 is less able to spread at high temperatures); however, the details were reformulated or selected in a way that made it misleading and untrue.

To answer the research questions regarding the most frequently used parts of speech in the given samples as well as the ability to create fake news patterns, our research revealed that although it is easy to distinguish the most common parts of speech, which in terms of the given fake news claims are nouns, verbs and prepositions, the attempt to create fake news patterns is much more challenging. Moreover, the comparative analysis of both samples does not provide us with satisfactory results. Nevertheless, our research analysis to some extent showed support for the first hypothesis H1, suggesting that the main fake news constructs change regarding real situations, i.e., there has been a shift in claims that appeared in March 2020 and November 2020. For instance, there is a growing number of misconceptions regarding vaccine development due to encouraging recent real news about coronavirus vaccines.

We cannot provide one specific model built on the basis of the corpus-based analysis of fake-news claims, thus our findings cannot support the second hypothesis, H2. However, by modelling the constructs of fake news and providing their frequency, we reason that both the emergence of the most popular claims and the presented grammatical patterns may help the readers to become more aware of how misconceptions about COVID-19 are structured and which constructs are most commonly reconfigured, selected or re-contextualised in ways that make them untrue.

## 6. Conclusions

Undoubtedly, fake news has a negative impact on individual users and the whole of society. The speed with which it is disseminated makes the process of its detection a great challenge. Believing that the presented study may contribute to a better understanding of the phenomena of fake news in the time of the COVID-19 pandemic, we are aware of its limitations. First, we shall start from the fact that the given analysis is neither comprehensive nor it is exhaustive, as



we investigate only a sample of fake news claims in English. It is thus possible that our findings may not be generalised to the general sharing of fake news on other subjects. Nevertheless, based on the findings of this study and their detailed analysis, we feel there is a strong need for Internet users to know how to confirm true information and its authenticity. Hence, giving them some tools and guidelines for in-depth investigation and examining evidence to confirm facts and dispel myths is the best way for researchers to combat the infodemic. We first introduced some fundamental concepts of disinformation, misinformation, and fake news. Then, we reviewed the literature concerning the phenomenon of fake news. In this research, we attempted to reveal that fake news mostly focuses on topics based on conspiracy theories and political actions. The items of fake news are mainly structured with the use of such verbs as ‘be’, ‘have’ ‘die’ and ‘spread’. Moreover, it has been shown that adjectives appear more frequently in fake news than in facts. Although the achieved results do not allow us to introduce a model of fake news which could be helpful to differentiate it from facts, our research has shown that it is possible to carefully monitor online misinformation and detect the most common constructs with which fake news occurs. Finding new trends and tendencies in the circulation of fake news and comparing them to the previous findings may have a positive impact on improving the classification accuracy and reducing the number of misconceptions. Finally, it could be said that the guidelines we offer may help to answer the recent call which suggests that Internet users should be more careful in news verification in a time of global (mis)communication.

## Appendix 1

Internet sources used for gathering data:

<https://www.who.int/docs/default-source/searo/thailand/12myths-final099bfbf976c54d5fa-3407a65b6d9fa9d.pdf>

<https://www.newsguardtech.com/covid-19-myths/>

<https://www.cdc.gov/coronavirus/2019-ncov/vaccines/about-vaccines/vaccine-myths.html>

<https://www.avert.org/coronavirus/covid-19-myths-and-facts>

<https://www.medicalnewstoday.com/articles/coronavirus-myths-explored>

<https://www.msn.com/en-ph/news/photos/coronavirus-fake-news-things-you-heard-about-covid-19-that-are-not-true/ss-BB10YA5F#image=6>

## References

- Allcott Hunt, Gentzkow, Matthew (2017): Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31 (2), 211–236.
- Antos Gerd (2019): Fake news. Dlaczego im ulegamy. Albo: „Urządź wam świat tak, jak mi się podoba”. *tekst i dyskurs – text und diskurs* 12, 191–213. Accessed November 11, 2021. DOI: 10.7311/tid.12.2019.11
- Apuke Oberiri Destiny, Omar Bahihay (2020): Fake news proliferation in Nigeria: Consequences, motivations, and prevention through awareness. *Humanities & Social Sciences Reviews* 8 (2), 318–327. Accessed November 18, 2020. <https://doi.org/10.18510/hssr.2020.8236>. Accessed November 18, 2020.
- Apuke Oberiri Destiny, Omar Bahihay (2020a): Fake news and COVID-19: modelling the predictors of fake news sharing among social media users. *Telematics and Informatics* [Article in press]. Accessed November 2, 2020. <https://doi.org/10.1016/j.tele.2020.101475>.
- Barua Zapan, Barua Sajib, Aktar Salma, Kabir Najma & Li Minzge (2020): Effects of misinformation on COVID-19 individual responses and recommendations for resilience of disastrous consequences of misinformation. *Progress in Disaster Science* 8 (2020) 100119. Accessed November 16, 2020. <https://doi.org/10.1016/j.pdisas.2020.100119>.
- Bastick Zach (2020): Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation. *Computers in Human Behavior*. [Journal Pre-proof] Accessed November 22, 2020. <https://doi.org/10.1016/j.chb.2020.106633>.
- Brennen J. Scott, Simon M. Felix, Howard N. Philip & Nielsen Rasmus Kleis (2020): *Types, Sources, and Claims of COVID-19 Misinformation*. Accessed November 20, 2020. <https://reuters-institute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation>.
- Duffy Andrew, Tandoc Edson, Ling Rich (2019): Too good to be true, too good not to share: the social utility of fake news. *Information Communication and Society* 1–15. Accessed November 16, 2020. <https://doi.org/10.1080/1369118X.2019.1623904>.
- Egelhofer Jana Lura, Lecheler Sophie (2019): Fake news as a two-dimensional phenomenon: a framework and research agenda. *Annals of the International Communication Association* 43 (2), 97–116. Accessed November 20, 2020. <https://doi.org/10.1080/23808985.2019.1602782>.
- Hou Zhiuan, Du Fanxing, Jiang Hao, Zhou Xinyu, Lin Lessa, Assessment, T., & Commission, N. H. (2020): Assessment of public attention, risk perception, emotional and behavioural responses to the COVID-19 outbreak: social media surveillance in China. *medRxiv* 2020.03.14.20035956. Accessed November 16, 2020. <https://doi.org/10.1101/2020.03.14.20035956>
- Howell Bethany (2020): *Infodemic: The Rise of Fake News During COVID-19*. Accessed November 16, 2020. <https://tech.co/news/fake-news-covid-19>.
- Huynh Duc Luu Toan (2020): The COVID-19 risk perception: A survey on socioeconomics and media attention. *Economics Bulletin* 40 (1), 758–764. Accessed November 16, 2020. <https://ideas.repec.org/a/ebl/ecbull/eb-20-00175.html>.
- Islam A. K. M. Najmul, Laato Samuli, Talukder Shamim, Sutinen Erkki (2020): Misinformation sharing and social media fatigue during COVID-19: An affordance and cognitive load perspective. *Technological Forecasting & Social Change* 159 (2020) 120201. Accessed November 20, 2020. <https://doi.org/10.1016/j.techfore.2020.120201>.
- Kaminska Izabella (2017): A module in fake news from the info-wars of ancient Rome. *Financial Times*. Accessed November 16, 2020. <https://www.ft.com/content/aaf2bb08-dca2-11e6-86ac-f253db7791c>

- Ku Y. L. Kelly, Kong Qiuyi, Song Yunya, Deng Lipeng, Kang Yi, Hu Aihua (2019): What predicts adolescents' critical thinking about real-life news? The roles of social media news consumption and news media literacy. *Thinking Skills and Creativity* 33, 100570. Accessed on November 18, 2020. <https://doi.org/10.1016/j.tsc.2019.05.004>.
- Lampos Vasileios, Maimuna S. Majumder, Elad Yom-Tov, Michael Edelstein, Simon Moura, Yoh-hei Hamada, Molebogeng X. Rangaka, Rachel A. McKendry, Ingemar J. Cox (2020): *Tracking COVID-19 using online search*. 93, 4–9. Accessed November 12, 2020. <https://arxiv.org/abs/2003.08086>.
- McGonagle, Tarlach (2017): “Fake news”: False fears or real concerns? *Netherlands Quarterly of Human Rights* 35 (4), 203–209. Accessed November 16, 2020. <https://doi.org/10.1177/0924051917738685>.
- Nougayrede Natalie (2018): In this age of propaganda, we must defend ourselves. Here's how. *The Guardian* (31/01/18) Accessed November 16, 2020. <https://www.theguardian.com/commentisfree/2018/jan/31/propaganda-defend-russia-technology>
- Pal Anjan, Banerjee Snehasish (2019): Understanding online falsehood from the perspective of social problem. In I. E. Chilua & S. A. Samoilenko (Eds.), *Advances in media, entertainment, and the arts (AMEA) book series. Handbook of research on deception, fake news, and misinformation online* (p. 1–17). Information Science Reference/IGI Global. Accessed November 18, 2020. <https://doi.org/10.4018/978-1-5225-8535-0.ch001>
- Pennycook Gordon, McPhetres Jonathon, Zhang Yunhao, Rand G. David (2020): Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy nudge intervention. *SAGE Journals. Psychological Science*, 31(7) 770–780. Accessed November 16, 2020. <https://doi.org/10.1177/0956797620939054>.
- Piller Ingrid, Zhang Jie, Li Jia (2020): Linguistic diversity in a time of crisis: Language challenges of the COVID-19 pandemic. *Multilingua* 2020, 39(5), 503-515. DOI: <https://doi.org/10.1515/multi-2020-0136>. Accessed October 23, 2020.
- Porter Jon (2020): Snopes Forced to Scale Back Fact-Checking In Face of Overwhelming COVID-19 Misinformation. *The Verge*, March 24, 2020, <https://www.theverge.com/2020/3/24/21192206/snopes-coronavirus-covid-19-misinformation-fact-checking-staff>. Accessed October 23, 2020.
- Posetti Julie, Matthews Alice (2018): A short guide to the history of 'fake news' and disinformation. A learning module for journalists and journalism educators. International Centre for Journalists. Accessed November 20, 2020. [https://www.icfj.org/sites/default/files/2018-07/A%20Short%20Guide%20to%20History%20of%20Fake%20News%20and%20Disinformation\\_ICFJ%20Final.pdf](https://www.icfj.org/sites/default/files/2018-07/A%20Short%20Guide%20to%20History%20of%20Fake%20News%20and%20Disinformation_ICFJ%20Final.pdf)
- Pulido M. Christina, Villarejo-Carballido Beatriz, Redondo-Sama Gisela, Gómez Alitor (2020): *COVID-19 infodemic: More retweets for science-based information on coronavirus than for false information*. *SAGE Journals. International Sociology* 35(4), 377-392. Accessed November 16, 2020. <https://doi.org/10.1177/0268580920914755>.
- Scheufele A. Dietram, Krause M. Nicole (2019): Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences of the United States of America* 116(16), 7662–9. Accessed November 20, 2020. <https://doi.org/10.1073/pnas.1805871115>.
- Tan Hua Kim, Woods Peter, Azman Hazita, Abdullah Imran Ho, Hashim Ruzy Suliza, Rahim Hajar Abdul, Idrus Mohd Muzhafar, Said Nur Ehsan Mohd, Lew Robert, Kosem Iztok (2020) COVID-19 insights and linguistic methods. 3L. *The Southeast Asian Journal of English Language Studies*, 26(2), 1-23. Accessed November 12, 2020. <http://doi.org/10.17576/3L-2020-2602-01>.
- Thompson Nik, Wang Xuequn, Daya Pratiq (2019): Determinants of news sharing behavior on social media. *Journal of Computer Information Systems*, 60(6), 593–601. Accessed November 1, 2020. doi:10.1080/08874417.2019.1566803.

- Vasu Norman, Benjamin Ang, Terri-Anne Teo, Shashi Jayakumar, Muhammad Faizal, and Juhi Ahuja (2018): FAKE NEWS: NATIONAL SECURITY IN THE POST-TRUTH ERA. Report. *S. Rajaratnam School of International Studies*, 2018. 5-9. Accessed November 18, 2020. <http://www.jstor.org/stable/resrep17648.5>.
- Wang Yuxi, Mckee Martin, Torbica Aleksandra, Stuckler David (2019): Social Science & Medicine Systematic Literature Review on the Spread of Health-related Misinformation on Social Media. *Social Science & Medicine* 240, 112552. Accessed November 18, 2020. <https://doi.org/10.1016/j.socscimed.2019.112552>.
- Wardle Claire, Derakhshan Hossein (2017): INFORMATION DISORDER: Toward an interdisciplinary framework for research and policy making. *Council of Europe report DGI(2017)09*. Accessed November 20, 2020. <https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-research/168076277c>.
- Waszak M. Przemyslaw, Kasprzycka-Waszak Wioleta, Kubanek Alicja (2018): The spread of medical fake news in social media – The pilot quantitative study. *Health Policy and Technology* 7 (2), 115–118. Accessed November 16, 2020. <https://doi.org/10.1016/j.hlpt.2018.03.002>.
- World Health Organisation (2020): Coronavirus disease (COVID-19) advice for the public: Myth-busters. Accessed November 1, 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters>.