

Beata Wójtowicz

University of Warsaw
ORCID 0000-0002-8819-3219

Kirusi or *virusi*? Corpus-based research on COVID-19-related terminology in Swahili¹

ABSTRACT

The most recent coronavirus outbreak that began in China in December 2019 has affected every aspect of our lives. Not only has it forced many changes in our daily routines but also influenced our language and brought new words and phrases that are constantly evolving to describe the new reality. In times of crisis, it is very important that people get reliable information in a language they know and understand. Therefore, many efforts have been made to provide information and educational materials to African people in their mother tongues or well-known local languages. This is also true for Swahili, which is the main vernacular language of East Africa. The beginning of the pandemic was accompanied by terminological chaos, as many new terms had to be invented to describe the new situation. In the case of Swahili, the Tanzanian National Kiswahili Council (BAKITA) published a terminology list with Swahili equivalents of English terms related to COVID-19. This list formed the basis of the author's study on the usage of such terminology, e.g., *UVIKO* and *virusi vya korona*, in Swahili news texts. The research was based on the corpus "Swahili News 2021" compiled for this purpose from online resources in the Sketch Engine corpus tool.

KEYWORDS: COVID-19, coronavirus pandemic, Swahili language, terminology development, terminology lists

¹ The research for this study was supported by the Polish National Science Centre grant no. 2018/31/B/HS2/01114.

1. INTRODUCTION

At the beginning of 2020, the world unexpectedly had to face a new challenge, an epidemic of a new variant of coronavirus that had already been observed in China at the end of the previous year (Chaplin 2020). At the time of writing this paper, in April 2022, we are still struggling with the virus. Fortunately, most of the restrictions imposed in 2020 have already been lifted in most countries, especially in Europe. People who have been exposed to the virus are no longer required to self-isolate, there is no mandatory (post-entry) quarantine, wearing masks is not obligatory anymore, and there are no restrictions on the size of public gatherings. However, it is still recommended to protect oneself and prevent the spread of COVID-19 by social distancing, washing one's hands frequently and wearing masks when maintaining a physical distance is not possible. One should also get vaccinated if possible and watch oneself for symptoms of the coronavirus. If any appear, one is advised to get tested, and if the result is positive, to stay home and self-isolate.

Daily business is running normally again, and schools have gone back to in-person teaching. Remote work and education have come to an end, even if some have found it quite convenient and are now calling for the introduction of hybrid learning. The blended mode is still used, especially in education and research, for meetings and conferences. In such cases, some students or conference participants attend the class/meeting in person, while others join in remotely.

It is no surprise that the current global coronavirus pandemic crisis has affected many spheres of our lives. For more than two years, we have had to learn how to deal with the new situation but also how to communicate about it. How to talk about the virus and describe its impact on our lives. Back in 2019, we were not yet familiar with such terms as “hybrid learning”, “self-isolation” or “social distancing”. Now, we all know what they mean and what is expected of us when someone refers to them.

This period has led to the creation of many new words and phrases, and changes have been observed in the frequency of use of terms that were already well established. Terms that were in use before COVID-19, though on a very small scale, include “pandemic”, “self-isolation”, “lockdown” and “quarantine”, among others. The term “coronavirus” was introduced by

virologists as early as the 1960s (Cierpich-Koziel 2020), but initially, its usage was limited to health-related contexts, discussed only in specialized journals. Today, it is one of the most frequently used terms, and language users have coined coronavirus-based neologisms that have become very popular, such as “coronaparty” (similar in meaning to “quaranteam”), “covidiot” or “coronacoaster”, a term used to describe the emotional experience of life during the pandemic. New terms are being invented all over the world by language users themselves, while at the same time, various organizations and governments are promoting more specialized terms and their uses.

The official name of the coronavirus that causes COVID-19 is SARS-CoV-2, which stands for “severe acute respiratory syndrome coronavirus 2”. It was introduced in February 2020 by the World Health Organization. COVID-19 means “coronavirus disease 2019”. The term “pandemic” was first used in June 2009 to describe the influenza pandemic of the time. Probably no one remembers it, but the term itself no longer surprises anyone since “pandemic” is among the most frequently used words in all the researched languages.

To facilitate understanding of the coronavirus outbreak and the spread of COVID-19 by individuals, as well as to enable and improve international communication, many glossaries and terminology lists have been created and distributed online. These include mono- and multilingual resources published by governments, dictionary publishers and various organizations involved in language documentation and development, health and translation services, etc.

The impact of the pandemic on languages has been studied from different perspectives since the outbreak of COVID-19. One approach is to look at corpus data and identify keywords and collocations that have been much more frequently used during the period in question than in the entire dataset of a given corpus. This will be discussed in later sections.

In general, the present paper deals with the coronavirus terminology in Swahili. It examines the recommendations of the Swahili language councils, terminology lists and other sources and verifies them against the practice of language use in East African and, more generally, Swahili online resources. The analysis was conducted on data extracted from Swahili news sources that are available online, with special attention paid to two newspapers, *Taifa Leo* from Kenya and *Habari Leo* from Tanzania.

2. OVERVIEW OF LITERATURE

In addition to media reports about the pandemic, the coronavirus crisis was reflected in scientific research. One very active field was linguistics. Linguists studying various languages collected vocabulary and data related to the coronavirus and created corpora of texts to analyze and interpret the pandemic-related discourse from a linguistic perspective. Some studies concentrated on language creativity (Lawson 2020), while others analyzed the discourse on the topic to see, on the one hand, how people's behavior had been modified by government and media messages and, on the other, how audiences perceived the COVID-19 pandemic (Yu et al. 2021). In her analysis of President Emmanuel Macron's speech broadcasted on French television on March 16, 2020, Frach (2021) noted that it used the rhetoric of fear and was full of sensational metaphors and references to war. The expression *nous sommes en guerre* 'we are at war', which appeared in it several times, was repeated around the world to describe the coronavirus crisis and evoke the expected public response. Similar strategies were used in communication concerning the threat in other languages. Both government and media messages were studied in different languages to identify ideologies hidden in the speeches and the use of metaphors (a comparative study by Gulzar et al. 2021; see also Bates 2020; Abdel-Raheem 2021; Waśko 2021; Przybysz and Wójtowicz 2022; on South Africa, see Rudwick et al. 2021).

The Oxford English Dictionary² has been tracking words and phrases that have become unexpectedly and overwhelmingly popular during the last two years. The data shows that the global discourse of the period was dominated by such words as "coronavirus" and "COVID-19". While "COVID-19" is a new term, "coronavirus" was previously used mainly in medical and scientific discourse, so the average language user was not familiar with it. Examining the different and changing contexts of the terms' usage reveals the perceptions and concerns of users at a given time.

² <https://public.oed.com/blog/corpus-analysis-of-the-language-of-covid-19/> [accessed 19.05.2022].

The frequency of words in Polish media is monitored at the University of Warsaw as part of the “Słowa na czasie” project.³ The loanword *koronawirus* ‘coronavirus’ entered the top ranking already in January 2020 and has been gaining popularity ever since. Cierpich-Kozieł (2020) conducted a study of *koronawirus* neologisms by examining media corpus data from that period.

Other research focused on the development of terminology and its standardization in different languages. Tomaszewska and Zawadzka-Paluckta (2020: 13) examined a parallel corpus of EU press releases in English and Polish to determine how multiword terms relating to the COVID-19 pandemic were translated from the former language into the latter in expert-to-lay communication. Not surprisingly, the terminology was not yet “fully stabilized” (Tomaszewska and Zawadzka-Paluckta 2020: 38) in 2020.

Much less attention has been paid to the terminology in languages other than English. Even in Europe, it was expected that translation services could be delayed due to the rapidly changing and evolving epidemiological situation, as the crisis posed a challenge in rendering COVID-19-related terminology. This also resulted in a greater terminological variation.

The threat of a language gap in coronavirus information was anticipated in Africa too. At the same time, the importance of disseminating information about COVID-19 in local languages was well recognized. Although European languages have the status of official languages in most African states, they are mainly accessible to people with a certain level of education, so local languages are used in everyday communication instead. Some of the media and educational institutions also provide news and information to local communities in the languages they know and use on a daily basis. Therefore, many initiatives were undertaken to standardize COVID-19-related terminology in African languages. Some of these were launched by various non-governmental organizations, bottom-up movements and volunteers; others by governmental bodies and language councils dedicated to the promotion and standardization of local languages.⁴ In the following section, I will focus on such initiatives concerning Swahili.

³ <http://www.slowanaczasie.uw.edu.pl/> [accessed 19.05.2022].

⁴ In Europe, one such example is the European Union’s multilingual database of terminology (Interactive Terminology for Europe 2020: <https://iate.europa.eu>), which has been updated

3. INFORMATION ON COVID-19 IN SWAHILI

During the coronavirus pandemic, a sudden influx of new terminology was recorded in all the world's languages. The need to provide crucial information in various local languages became urgent, as it was observed that only information in languages that people understood, i.e., their mother tongues, could save lives. As Ahmad (2020) put it, “[t]he availability of knowledge in the languages of the people not only ensures that the message is not misunderstood due to a poor level of command of the official language but also creates trust, which increases the likelihood of acceptance”. Furthermore, Rudwick et al. (2021: 243) noted that “South Africans can be expected to be most receptive and responsive to briefings and information of epidemiologists, health and government officials when the message is conveyed in their first and home language”. Also, Translators without Borders found that “providing health information in Swahili ... produced a significant increase in comprehension compared with providing the same information in English. ... the local form of Swahili was the most effective language for risk communication and community engagement for the Ebola response in Goma, Democratic Republic of the Congo, compared with the French and standard Swahili languages” (Utunen et al. 2020).

At the same time, most African languages are under-resourced, endangered and hinder access to globally available information and services. Terminological confusion could be observed in many languages, which prompted the need for terminological standardization, especially in the field of medical language translation (Svongoro et al. 2021).

Many bottom-up initiatives were undertaken by local language users themselves. Volunteers for WikiAfrica translated COVID-19-related online content into nearly 20 African languages and created many new words to describe the situation (Harrisberg and Eaton 2021). The Engage Africa Foundation⁵ published COVID-19 public health guidelines in 18 of the most widely spoken African languages.

to register and recommend new COVID-19-related terms in the official languages of the EU.

⁵ <https://www.engageafricafoundation.org/covid-19?lang=swh#infographics> [accessed 6.07.2022].

Swahili stands out from other African languages. It is a well-established, well-resourced (compared to others) and well-studied language, spoken by millions of people in East Africa. Its importance and usefulness in communicating with the local population was recognized already during the colonial period when efforts were made to standardize the language. Since then, various governmental and scientific institutions have stimulated and regulated its development. Swahili is widely used in everyday communication, administration, media and education. Swahili scholars have demonstrated that an African language can be successfully used in scientific research and academic writing, therefore a lot of attention has been paid to terminology development and translation studies.

The Interterritorial Language Committee was established by British colonial rule back in the 1930s with the task of language standardization. After fulfilling its role, the Committee was reorganized in 1964 into a research institution called Taasisi ya Uchunguzi wa Kiswahili (TATAKI) and then integrated as a unit of the University of Dar es Salaam in 1970. Meanwhile, Baraza la Kiswahili la Taifa (BAKITA) ‘The National Kiswahili Council’ was founded in Tanzania in 1967 to promote, standardize and further develop the language. Concurrently, other regional and national councils, such as BAKIZA (Zanzibar), BAKIKE (Kenya) and BAKIU (Uganda), were established in countries where Swahili was promoted as a national and later official language.

From the very beginning, information about the coronavirus outbreak was widely published in Swahili by various African and non-African health and (non-)governmental agencies, and the situation was commented on by Swahili-language mass media. In line with the assumption that “the best way to fight its [COVID-19] outbreak is to make people aware of how it spreads and how it can be stopped” (Ahmad 2020), various information channels were used to spread the news. Many websites promoted prevention resources and educational materials in the form of leaflets, instructional videos and even children’s books, such as *Hakuna Jinamizi Nje: ni Kirusi* (Ndibalema et al. 2020).

According to a study that tested knowledge, attitudes and practices regarding the coronavirus outbreak among Swahili speakers at the beginning of the crisis in Tanzania, the educational and information

materials were well distributed and reached their intended audience. The report stated that “findings revealed good knowledge, optimistic attitudes, and appropriate practices towards preventing COVID-19 infection. Suggesting that community-based health education programs about COVID-19 is helpful and necessary to control the disease” (Rugarabamu et al. 2020: 3).

3.1. COVID-19-RELATED TERMS AND TERMINOLOGY LISTS IN SWAHILI

Swahili is spoken by approximately 100 million users in the vast region of East Africa (Eberhard et al. 2021). It has many regional variants, and the large area of its use hinders its standardization. In the face of the new coronavirus reality, it was necessary to come up with new terms to name it. The pandemic initially caused terminological chaos. English borrowings for coronavirus and COVID-19 were commonly used at first, but with time, several proposals for naming the phenomenon in Swahili were introduced. The newly-created names were mainly based on two words meaning disease/sickness in Swahili, which are *ugonjwa* and *maradhi*. Suggestions included such acronyms as *MAVIKU-19*, which stands for *Maradhi ya Virusi vya Korona* ‘coronavirus disease’, *UVIKU-19* for *Ugonjwa wa Virusi vya Korona* ‘coronavirus disease’ and *VVK-19* for *Virusi vya Korona* ‘coronavirus’.

Among the most important proposals was the acronym *UVIKO 19* – *Ugonjwa wa Virusi vya KORona*, an equivalent of COVID-19 introduced by BAKITA (2021). The Tanzanian Council also suggested adapting the Swahili spelling to the pronunciation by replacing the letter “c” with the letter “k” in other terms. As a result, *korona* was proposed instead of *corona*, and the name for the coronavirus became therefore *virusi vya korona*.

Before BAKITA published its recommendations, an interesting discussion about the equivalent of the English word “virus” was initiated by Swahili scholars. They debated whether *virusi* was the plural form of *kirusi*, formed according to the rules of the *ki-/vi-* noun class, where the prefix *ki-* signals the singularity and *vi-* the plurality of a noun, as in *kiti* ‘chair’ – *viti* ‘chairs’, or whether it was a generic name for a virus with no singular-plural distinction and should be used in this form in all cases while adhering to the agreement rules of the *vi-* class. The opinions of the researchers varied,

but the variant proposed by BAKITA and the data presented further in this article support the second option.

Many borrowings from English can also be found among other COVID-19-related terms, such as *maski* ‘mask’ and *glavu* ‘gloves’. However, other strategies for creating the terms have also been used. These include semantic extension, when new meanings are assigned to existing words, e.g., *barakoa* or *kizuzui*⁶ ‘veil’ for “mask” and *janga* ‘calamity’ for “epidemic/pandemic”, as well as derivation or paraphrasing, e.g., *kuepuka msongamano/kutokaribiana* lit. ‘avoiding crowds/not getting close to each other’ for “social distancing” and *zuio la kutotoka nyumbani* lit. ‘a ban on leaving the house’ for “lockdown”.

A detailed review of the term-creation strategies is beyond the scope of this paper, therefore I will now focus on analyzing the use of the terms in different sources.

The importance of establishing a standard vocabulary concerning COVID-19 so that information could be communicated with clarity and precision was recognized by BAKITA, which in December 2020 published online *Istilahi za Korona* ‘Coronavirus Terminology’. It was then updated in 2021 as part of the document *Istilahi sanifu za hali ya hewa, UVIKO-19 na mazingira* ‘Standard Terminology – Climate, COVID-19 and Environment’ (BAKITA 2021). In the introduction, it was stated that “[t]hese terms have been a great help to various stakeholders as they have closed the existing gap of shortage of terminological synonyms in the relevant fields”⁷ (BAKITA 2021: iv). The list of terms in the PDF file includes Swahili translations for 103 English scientific terms relating to COVID-19. As I have mentioned earlier, BAKITA suggested to call the SARS-CoV-2 coronavirus *UVIKO-19 (Ugonjwa wa Virusi vya Korona* ‘coronavirus disease’) and to use the letter “k” instead of the English “c” in *korona*.

⁶ I would like to thank Mohamed R. Karama for the information on this term. As he explained in an email to me, *kizuzui* comes from the face covering of Swahili women with the now archaic/extinct *buibui la wazee* ‘old women veil’. It is a way of covering with a small cloth attached to the *buibui* that covers the nose and part of the face but leaves an opening for the eyes.

⁷ “[i]stilahi hizi zimekuwa ni msaa mkuu kwa wadau mbalimbali kwani zimeziba pengo lililokuwapo la uhaba wa visawe vya istilahi katika nyuga zinazohusika”.

Another initiative was undertaken by Oxford University Press – Oxford Languages. The “Covid 19 Language Hub”⁸ project provides key COVID-19 terminology translated into the world’s major languages, among them African, including Afrikaans, Northern Sotho, Setswana, Xhosa, Zulu and Swahili. The list in the PDF file contains 73 Swahili translations of English scientific terms, but also words and phrases relating to government and individual actions to stop the spread of the virus.

“TICO 19: the Translation Initiative for COvid 19”⁹ was founded by several academic and industry organizations (Amazon, Appen, Facebook, Google, Microsoft, Translated) that collaborated with Translators without Borders to prepare materials on COVID-19 in more than 100 world languages to be used by professional translators and for training state-of-the-art Machine Translation (MT) models. It recorded 328 English terms with Swahili translations and made them available for download in a CSV format. The translation benchmark was created by combining English open-source data from various sources. It is not a simple wordlist and includes many detailed phrases and collocations that can be useful in translation, e.g., ‘14 days in isolation’ *kujitenga kwa siku 14*, ‘2019 novel coronavirus’ *virusi vipya vya korona 2019*, ‘about coronavirus’ *kuhusu virusi vya korona*, ‘avoid exposure’ *epuka kujiweka hatarini*.

Translators without Borders¹⁰ also published a terminology list that contains 185 English terms with Swahili equivalents. It was created to identify “key terminology that people use to talk about COVID-19 as well as commonly used technical terms to develop a multilingual, plain-language glossary”.

In general, all the above-mentioned lists contain a lot of vocabulary. Unfortunately, they differ significantly from one another. A comparison of the list published by BAKITA with the one compiled by the Oxford Languages (OL) shows that there are only 20 identical English terms and 10 identical Swahili equivalents in both of them (see example [1]). The non-equivalent translations differ in lexical choices, as shown in example (2).

⁸ <https://languages.oup.com/covid-19-language-resources/> [accessed 5.05.2022].

⁹ <https://tico-19.github.io/> [accessed 5.05.2022].

¹⁰ <https://translatorswithoutborders.org/twb-glossary-for-covid-19> [accessed 5.05.2022].

- (1) Equivalent translations
 - coronavirus *virusi vya korona*
 - work from home *fanyia kazi nyumbani*
- (2) Non-equivalent translations
 - face covering
 - vikinga uso* (BAKITA), *kufunika uso* (OL)
 - herd immunity
 - kingajamii* (BAKITA), *kinga ya kundi* (OL)
 - lockdown
 - zuio la kutotoka nyumbani* (BAKITA),
 - kusitisha safari za ndani na nje ya eneo fulani* (OL)

Based on this data, we can further analyze the use of terminology and track variability in the use of particular terms. To deepen the study, I collected more data using a corpus management tool, i.e., Sketch Engine.

4. RESEARCH METHODS

The media play a vital role in influencing citizens' attitudes and behavior since the public turns to them for up-to-date information. Citizens also rely heavily on the media in crisis situations. From the very beginning of the coronavirus outbreak, the media have constantly monitored and reported on the situation, which is why the news corpus seems to be an appropriate source of data for the analysis of COVID-19-related terms.

Based on a self-built corpus called "Swahili News 2021", this study aims to explore how the COVID-19-related terminology in Swahili was used in various information sources. For this purpose, Swahili news and communications containing the keywords *coronavirus*, *koronavirus*, *COVID*, *UVIKO*, *mgonjwa*, *virusi*, *chanjo*, *maradhi* and *hospitali* were collected and aggregated in text format in Sketch Engine.¹¹ Manual filtering of press and media releases based on the above keywords resulted in 6,000 documents from various online news sources, such as *BBC*, *Deutsche Welle*, *Taiifa Leo*, *Mtanzania*, *Mwanahalisi online*, *Mwanaspoti*, *Habari Leo*, *Mwananchi* and *Jamii Forums*. The entire text corpus consists of over 5 million words.

¹¹ <https://www.sketchengine.eu/> [accessed 5.05.2022].

The “Swahili web 2014” (SwahiliWaC) already available in Sketch Engine was used as a reference corpus. It was crawled from various online sources in 2014 and consists of 18 million words.

To identify relevant terms, the frequency lists of the newly compiled and the reference corpora were compared first, and then the keywords and term extraction feature of Sketch Engine was used. Keywords – both individual words and multiword expressions – indicate which topics are characteristic of a given corpus. The Sketch Engine tool compares the corpora and identifies what is unique or typical in the analyzed corpus compared to the reference corpus, presenting only those elements that appear more frequently in the former.

In the second step, two sub-corpora were built in order to track the differences in term usage in Tanzanian and Kenyan varieties of Swahili. One was created using texts from the Tanzanian newspaper *Habari Leo* and the other from the Kenyan newspaper *Taiifa Leo*. The analysis was based on a comparison of the newspapers’ wordlists and the results of keyword identification.

5. RESULTS AND DISCUSSION

Frequency wordlists contain those words that occur most frequently in the analyzed data and, together with keywords extraction, can indicate the most relevant topics. Due to the nature of the compiled corpus, which was intentionally built from coronavirus-related texts, the frequency of COVID-19-related terms was expected to be high, probably higher than in the general corpus from the same period. Nevertheless, the main aim of the study was to identify terms that were frequently used to describe the outbreak and compare them with those proposed by BAKITA (2021) to evaluate their usage.

The first 500 items on the “Swahili News 2021” frequency wordlist were analyzed to identify terms related to COVID-19. Their context of use was then checked using the concordance tool. For comparison, the terms were searched for among the 1,000 most frequent words in the reference corpus to see how the frequency of use of generic words such as “hospital”

or “symptom” had changed because of the pandemic. It turned out to be, respectively, 3 and 4 times higher than before. The frequency of use of *ugonjwa* ‘disease’ had increased 10 times. Some of the terms that were known and used before the crisis, e.g., *maambukizi* ‘infection’ or *chanjo* ‘vaccination’, were not found among the top 1,000 items on the Web 2014 list at all.

Table 1 below shows how often various COVID-19-related terms were used in the “Swahili News 2021” corpus (News 2021) compared to the corpus compiled in 2014 (Web 2014). The numbers in the columns indicate the term’s place on the wordlist in the given corpus; the lower the number, the more frequently the term was used. The lack of a number indicates the absence of the term in the analyzed subset of the frequency wordlist.

Table 1. Comparison of the frequency wordlists from the two analyzed corpora

News 2021	Lemma	Web 2014
62	<i>afya</i> ‘health’	382
75	<i>chanjo</i> ‘vaccination’	–
77	<i>ugonjwa</i> ‘illness, disease’	783
89	<i>virusi</i> ‘virus’	–
161	<i>magonjwa</i> ‘sickness’	–
196	<i>maambukizi</i> ‘infection’	–
197	<i>dalili</i> ‘symptom’	879
227	<i>corona</i>	–
270	<i>hospitali</i> ‘hospital’	796
278	<i>COVID-19</i>	–
460	<i>janga</i> ‘calamity; epidemic’	–

The data presented in Table 1 indicate that COVID-19-related terms are well represented in the recent news corpus. The frequency of their use is so high that they can be expected to be found as entries even in very small dictionaries. So far, the lemma *virusi* (nor *kirusi*) has not been recorded in comprehensive dictionaries of Swahili, such as those compiled by TUKI (2001) and BAKIZA (2010), even though it was widely used before the current crisis as part of the expression *virusi vya UKIMWI* ‘AIDS virus’. The Web 2014 corpus contains nearly 1,500 occurrences of *virusi*. Moreover, some

generic words that were frequently used even before the coronavirus crisis, e.g., *hospitali* ‘hospital’ and *dalili* ‘symptoms’, have now gained in popularity.

Judging by the keywords of the News 2021 corpus, coronavirus-related topics constitute an important part of it. The top 50 keywords include such terms as *corona*, *COVID*, *COVID-19*, *barakoa* ‘mask’, *korona* ‘corona’, *UVIKO* ‘COVID’, *coronavirus*, *UVIKO 19* ‘COVID-19’, *chanjo* ‘vaccination’, *karantini* ‘quarantine’, vaccine, Wuhan. The terms proposed by BAKITA (2021), i.e., the Swahili equivalent for COVID – *UVIKO* and *korona* ‘corona’, are among the top 10 keywords.

The words accompanying the terms and the changing contexts in which they occur provide information on users’ perceptions and concerns. Table 2 below shows the top 20 collocates of *korona/virusi vya korona* ‘corona(virus)’ in the News 2021 corpus, that is, words occurring near the term with a statistically significant frequency.¹² It appeared most often together with such nouns as “outbreak”, “spreading” and “pandemic” and verbs like “spread”, “kill” and “infect”, which suggests that the texts portrayed it as a highly contagious virus that must be fought against because it threatens people’s health and even kills them.

Table 2. Top 20 collocates of *korona/virusi vya korona* ‘corona(virus)’ in the News 2021 corpus

Noun	English translation	Verb	English translation
<i>kipimo</i>	health examination	<i>sambaa</i>	spread
<i>kanga</i>	pandemic	<i>sababisha</i>	cause
<i>ueneaji</i>	spreading	<i>pima</i>	examine
<i>mlipuko</i>	outbreak	<i>ambukiza</i>	infect
<i>chanjo</i>	vaccination	<i>enea</i>	spread
<i>maambukizi</i>	infection	<i>ua</i>	kill
<i>ugonjwa</i>	sickness	<i>athiri</i>	infect
<i>dalili</i>	symptoms	<i>pasa</i>	be obliged
<i>aina</i>	kinds	<i>chanja</i>	vaccinate
<i>wagonjwa</i>	sick people	<i>pambana</i>	fight with

¹² The collocates are presented separately for nouns and verbs, as analyzed in Sketch Engine.

The list of the top 500 n-grams (multiword expressions) from the News 2021 corpus also includes many COVID-19-related phrases. These are, for example, *virusi vya corona* ‘coronavirus’, *chanjo dhidi ya virusi* ‘vaccination against the virus’, *ugonjwa wa corona* ‘coronavirus disease’, *janga la corona* ‘coronavirus pandemic’, *maambukizi ya virusi* ‘virus infection’, *mfumo wa kinga* ‘immune system’, *dalili za covid 19* ‘COVID-19 symptoms’. N-grams help to identify discursive units of language that can be considered fixed phrases and multiword expressions and used in further linguistic analysis of a language or language teaching.

A cursory analysis of the frequency wordlists of two newspapers, Kenyan *Taifa Leo* and Tanzanian *Habari Leo*, shows that coronavirus-related terms were used more often in Tanzanian texts, despite the fact that both countries’ attitudes toward the virus might indicate otherwise.¹³

To check whether the terms used in the newspapers corresponded to those proposed by BAKITA (2021), the occurrences of selected terms – including COVID, *UVIKO*, *kirusi*, *virusi*, corona and its Swahili variant *korona* – were further analyzed.

In the Kenyan *Taifa Leo*, the term COVID was used 12 times, while no occurrences of *UVIKO* were noted. Corona was used 38 times, in phrases such as *virusi vya corona* ‘coronavirus’, *chanjo ya corona* ‘coronavirus vaccination’, *dhidi ya corona* ‘against coronavirus’, *kudhibiti corona* ‘to control coronavirus’, *janga la corona* ‘coronavirus pandemic’, *hali ya corona* ‘coronavirus condition’, *maambukizi ya corona* ‘coronavirus infection’, *ugonjwa wa corona* ‘coronavirus disease’. There were no instances of the Swahili variant for corona, i.e., *korona*, and no occurrences of *kirusi* ‘virus’, while 11 examples were noted for *virusi*.

In the Tanzanian *Habari Leo*, the term COVID was used 17 times and *UVIKO* 16 times. Corona had 26 occurrences, in similar phrases as in *Taifa Leo*. As in the former newspaper, no instances of *korona* and *kirusi* were noted. *Virusi* was used 25 times.

The data is scarce, but it clearly shows that Tanzanian Swahili was more responsive to the language recommendations announced by the National Kiswahili Council (which is Tanzanian).

¹³ From the beginning of the outbreak, Tanzania’s then-president John Magufuli had been one of the most prominent coronavirus skeptics and had even stopped testing for the virus. Also, already in May 2020, he declared the country free of the virus (Nicolini 2021).

There are many more occurrences of the search terms throughout the News 2021 corpus. COVID appears 1,300 times and corona over 3,000 times. Swahili variants of these terms, namely, *UVIKO* and *korona*, have 600 and 400 occurrences, respectively. Further analysis of the sources where *UVIKO* and *korona* have been noted shows that these terms are used by non-African newspapers, like *BBC*, *DW*, *Vatican News*; Tanzanian newspapers, e.g., *Mwanaspoti*; and Tanzanian private webpages and forums, such as *Muungwana Blog ya Kitanzania*, *Jamii Forums*, etc.

6. CONCLUSION

The COVID-19 crisis has affected people and their lives around the world, also impacting languages and the way we communicate about it. During the pandemic, providing information about the health crisis in languages that people use on a regular basis became a necessity. It has been proven that reliable information that is easy to understand by everyone can save people's lives. That is why it was so important to produce information about the coronavirus pandemic in various local languages around the world.

The new situation caused terminological chaos and led to a proliferation of new words and phrases. Some of the terms appeared in the dictionaries for the first time, others gained new meanings and frequency of use. All languages faced the same problem of how to talk and inform about the ongoing crisis in a way that everyone could understand. English became the source of new terms that had to be introduced into other languages. Many initiatives were undertaken to develop and standardize terminology. One of these was the COVID-19-related terminology list published by the Tanzanian National Kiswahili Council (BAKITA 2021), which proposed Swahili equivalents for English terms. Among other things, it suggested using *UVIKO-19* as the equivalent of COVID-19 and standardizing the name for "coronavirus" as *virusi vya korona*.

Drawing on corpus data compiled in 2021, the article examined the extent to which the terminology proposed by BAKITA has been used in Swahili media texts. The terms *UVIKO* and *korona* were used thousands of times in the analyzed corpus; in fact, they were among the top 10 keywords.

They appeared mainly in non-African newspapers and Tanzanian private webpages and forums. The selected terms were also searched for in two newspapers, one Kenyan and the other Tanzanian. It was noted that neither of them used words such as *kirusi* and *korona*. The Kenyan *Taifa Leo* also did not use the term *UVIKO* instead of COVID, whereas the Tanzanian *Habari Leo* did so several times.

In summary, the terminological chaos has not yet been resolved for Swahili, but efforts have been made in that direction. Recommendations published by the Tanzanian National Kiswahili Council have reached a wide audience and are being implemented, although mainly by Tanzanian and non-African bodies. Kenyan Swahili is so far more influenced by English borrowings.

REFERENCES

- Abdel-Raheem, Ahmed. 2021. "Reality Bites. How the Pandemic Has Begun to Shape the Way We, Metaphorically, See the World". *Discourse & Society* 32/5: 519–541.
- Ahmad, Rizwan. 2020. "Multilingual Resources Key to Fighting COVID-19". *Language on the Move*. Online publication [accessed 5.05.2022].
- BAKITA (Baraza la Kiswahili la Taifa). 2021. *Istilahi sanifu za hali ya hewa, UVIKO-19 na mazingira*. Online publication [accessed 5.05.2022].
- BAKIZA (Baraza la Kiswahili la Zanzibar). 2010. *Kamusi la Kiswahili Fasaha*. Nairobi: Oxford University Press.
- Bates, Benjamin R. 2020. "The (In)Appropriateness of the WAR Metaphor in Response to SARS-CoV-2. A Rapid Analysis of Donald J. Trump's Rhetoric". *Frontiers in Communication* 5/50: 1–12.
- Chaplin, Steve. 2020. "COVID-19. A Brief History and Treatments in Development". *Prescriber* 31/5: 23–28.
- Cierpich-Kozieł, Agnieszka. 2020. "Koronarzczywistość – o nowych złożeniach z członem korona- w dobie pandemii". *Język Polski* 100/4: 102–117.
- Eberhard, David M., Gary F. Simons and Charles D. Fennig (Eds.). 2021. *Ethnologue. Languages of the World*. 24th edition. Dallas, TX: SIL International. <http://www.ethnologue.com> [accessed 5.05.2022].
- Frach, Sylwia. 2021. "Retoryka strachu w czasie pandemii Covid-19. Przypadek Emmanuela Macrona". *Athenaeum. Polskie Studia Politologiczne* 70/2: 7–19.
- Gulzar, Kanwal, Saira Asghar Khan, Imdad Ullah Khan, Saood Khan and Sidra Amina. 2021. "An Analysis of War Language in COVID-19 Speeches. A Comparative Study". *Humanities & Social Sciences Reviews* 9/3: 1399–1412.

- Harrisberg, Kim and Kristi Eaton. 2021. "Young Africans Go Online to Preserve Local Languages, Fight COVID-19". *Thomson Reuters Foundation News*. Online article [accessed 5.05.2022].
- Lawson, Robert. 2020. "Coronavirus Has Led to an Explosion of New Words and Phrases – and that Helps Us Cope". *The Conversation*. Online publication [accessed 5.05.2022].
- Ndibalema, Placidius, Fredrick M. Nafukho, Machuma A. Muyia, Beverly J. Irby, Rafael Lara-Alecia, Nahad Abdelrahman, Fuhui Tong and Pam Schiller. 2020. *Hakuna Jinamizi Nje: ni Kirusi*. College Station, TX: Texas A&M University. Online publication [accessed 5.05.2022].
- Nicolini, Cristina. 2021. "From VVU/UKIMWI (HIV/AIDS) to UVIKO-19 (COVID-19). An Epistemological Analysis of Pandemics in Tanzania through Swahili Literature". *Kervan. International Journal of Afro-Asiatic Studies* 25/2: 63–88.
- Przybysz, Zuzanna and Beata Wójtowicz. 2022. "Koronawirus w Sejmie – jak posłowie mówili o pandemii?" In: Jolanta Łącka-Badura, Anna Iwańska and Elwira Poleszczuk (Eds.), *Języki specjalistyczne w badaniach i praktyce 2*. Katowice: Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach. 231–246.
- Rudwick, Stephanie, Zameka Sijadu and Irina Turner. 2021. "Politics of Language in COVID-19. Multilingual Perspectives from South Africa". *Politikon* 48/2: 242–259.
- Rugarabamu, Sima, Mariam Ibrahim and Aisha Byanaku. 2020. "Knowledge, Attitudes, and Practices (KAP) towards COVID-19. A Quick Online Cross-Sectional Survey among Tanzanian Residents". Online report [accessed 5.05.2022].
- Svongoro, Paul, Monwabisi Ralarala and Raymond Damba. 2021. "Face-to-face with COVID-19 Untranslatability in African Languages. Experiences of English–Shona Translators in Zimbabwe". *South African Journal of African Languages* 41/3: 249–258.
- Tomaszewska, Aleksandra and Natalia Zawadzka-Paluckta. 2020. "Translating a Pandemic. A Corpus Study of COVID-19 Multi-Word Terminology in EU Press Releases". *Beyond Philology* 17/4: 11–44.
- TUKI (Taasisi ya Uchunguzi wa Kiswahili). 2001. *Swahili-English Dictionary*. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili, Chuo Kikuu cha Dar es Salaam.
- Utunen, Heini, Ngouille Ndiaye, Corentin Piroux, Richelle George, Melissa Attias and Gaya Gamhewage. 2020. "Global Reach of an Online COVID-19 Course in Multiple Languages on OpenWHO in the First Quarter of 2020. Analysis of Platform Use Data". *Journal of Medical Internet Research* 22/4: e19076, doi: 10.2196/19076
- Waško, Jakub. 2021. "Analiza porównawcza informacji o początku epidemii koronawirusa w Polsce przedstawianych w największych portalach informacyjnych w kraju". *Świat i Słowo. World and Word* 36/1: 250–262.
- Yu, Hangyan, Lu Huiling and Jie Hu. 2021. "A Corpus-Based Critical Discourse Analysis of News Reports on the COVID-19 Pandemic in China and the UK". *International Journal of English Linguistics* 11/2: 36–45.