

Preliminary Findings of an Analysis of the COVID-19 English Terminology Used in Electronic News Media in Sri Lanka

W Abeyawickrama[#] and K Amaratunga

Faculty of Management, Social Sciences and Humanities, General Sir John Kotelawala Defence University, Sri Lanka

[#]wimansha@kdu.ac.lk

Abstract— COVID-19 outbreak has brought several language changes mainly in terminology. Some existing analyses on English lexicon usage during the pandemic highlight the frequent use of words which were infrequent before the pandemic and also new coinage. We used a corpus approach to conduct a preliminary analysis of COVID-19 related vocabulary used in electronic news media in Sri Lanka and identified that some health-related terminology used in many other contexts can also be seen in the Sri Lankan context. Apart from that, some localized COVID-19 related terminology has been found in the corpus. The collocation analysis of the words Coronavirus and COVID-19 shows that Coronavirus is used as a noun and a noun modifier while COVID-19 is mostly used as a noun modifier in the corpus.

Keywords— *COVID-19, coronavirus, corpus, collocations, frequency*

I. INTRODUCTION

COVID-19 pandemic has brought changes to almost all sectors in the world and introduced new norms, regulations and procedures. Changes that occurred in language use is second to none. COVID-19 has been a special situation where new vocabulary emerged rapidly and they became common words in society within a very short time. On 11 February 2020, Oxford English Dictionary added the word 'Covid-19' as a new entry and recognized it as a noun (OED, 2020). Since then, the world came across several COVID-19 terms in English and equivalents in other languages. When it comes to English vocabulary that emerged during the pandemic, it is possible to recognize two groups of words: ones which had already been there, but not commonly used in public and newly coined words. Most words used during the pandemic seem to belong to the former group and a very few to the latter group.

There have been several studies on the emergence and use of English vocabulary during the pandemic in several countries; however, to our knowledge, no

study has investigated pandemic related new English terminology used in Sri Lanka, and apparently no such study has so far been undertaken on pandemic related Sinhala (native language of the majority in Sri Lanka) terminology either.

The aim of the overall study is to analyse the COVID-19 terminology that emerged in Sri Lanka in both English and Sinhala in various news media and social media. In this paper, we discuss only the preliminary findings of this ongoing study and present the frequencies of some commonly used words in English news media in Sri Lanka within the first wave of the pandemic, that is from March 2020 to October 2020. We also discuss some frequent collocations that emerged in the corpus.

II. LITERATURE REVIEW

Language change takes place over centuries and as a result language structures and vocabulary change. There can be new borrowings, coinage, meaning changes and structural changes and this evolution is a natural process that any language undergoes. Language change can also occur over short spans of time, which has been identified as *current change* (Aarts, et al., 2013). Current change also refers to changes that took place in a language in the recent past.

COVID-19 has affected the whole world and there has been no similar situation in the recent history that has affected the world in the same way. With the spreading of COVID-19, the world also came across a large number of new terms; some are medical terms which were already there but uncommon and some terms have been coined. These terms became viral mainly due to social media and entered the linguistic systems of almost all languages. For examples, words such as *coronavirus, social distancing, flattening the curve, self-isolation, quarantine, lockdown, hotspots* and *travel restrictions* are some such words and phrases that have been regularly used in media for the past one and half years. As Upadhyay and Maroof

(2020) highlight, some of these words have been in the linguistic systems for centuries although they may not have been in public use. They also emphasize that these terms became widespread due to the more technologically connected world that exists at present.

Ezzarrouki (2020) compiled the health care terminology used during COVID-19 pandemic and identifies that coronavirus is a word that has been overwhelmingly used since the outbreak. Several collocates of coronavirus such as *Coronavirus outbreak*, *novel Coronavirus*, *spread of Coronavirus* and *fight the Coronavirus* emerged simultaneously. *Infodemic* has also been identified as a common term used during the pandemic. Newly emerged immunity lexicon includes words and phrases such as *healthy life style*, *normal hygiene* and *physical distancing* and words and phrases such as *social distancing*, *front-line health care workers*, *working from home* and *social distancing* have also been recognized as new terminology that emerged (ibid). Ezzarrouki (2020) recognizes *confirmed cases*, *viral shedding*, *the viral load*, *the virus incubation period*, *quarantine*, *self-quarantine*, *self-isolation*, *home-isolation*, *lockdown* and *stay-home* as some of the quarantine lexicon.

There have been some analyses on how COVID-19 terminology is used in electronic news media, social media and on the Internet. Xue et al. (2020) identified 11 emerging themes in their analysis of 1.9 million Tweets written in English, and *health emergency*, *corona virus*, *fake news*, *self-isolate*, *covid 19*, *number people*, *breaking news*, and *health officials* were some of the words that emerged in most number of Tweets. Valentin et al.'s (2020) corpus study of news analysis identified a shift of vocabulary used from early stages of the pandemic to later stages: words related to *pneumonia symptoms* and *mystery illness* were common in the early stage, but vocabulary related to *virus family* and *specific COVID-19 acronyms* have been common in the later stages. Bowker (2020) studied French vocabulary used during the pandemic in Canada and Europe and emphasises that the terminology considerably overlaps; however, there is also localization in order to ensure effective communication of important information to target populations.

Not only the emerging vocabulary, but also the message conveyed through COVID-19 terminology has been the focus of some of the studies. For example, Fan et al. (2020) analysed over 3 million Twitter posts at early stages of the pandemic in order to understand social stigmatization on China and

Chinese people. They found out that while social media plays an important role in raising public awareness in a crisis situation like the pandemic, it also substantially spread terminology associated with hate speech during the pandemic. In another study, Karami and Anderson (2020) identified 11 topics relating to quarantine related issues in their analysis of 80,000 Tweets with anti-quarantine comments.

These existing studies highlight that research on COVID-19 vocabulary takes two paths: one is on the new terms emerging and the other is a deeper analysis of terminology used to convey messages during the pandemic. While recognizing the importance of both strands, the current study has been designed to (1) analyse the English terminology that emerged in Sri Lankan news media and social media and to (2) analyse how such terms have been used to convey meaning. In this paper, we present some preliminary findings related to the above number (1) in which we analysed the terminology used in electronic news media in Sri Lanka during the first wave of the pandemic. We attempt to answer the following research question.

What are the frequently occurred COVID-19 pandemic related English words in electronic news media in Sri Lanka during the first wave of the pandemic?

III. METHODOLOGY

A. Procedure

We used a corpus approach in this analysis and created a corpus of all English electronic news media reports in Sri Lanka published on the internet between 1 March 2020 and 30 September 2020. *Sketchengine*, an online corpus tool dedicated for educational purposes, was used to build the corpus and also to analyse the data. When building the corpus, we used key words *Covid* or *Coronavirus* and *Sri Lanka* in order to choose and categorize the news items available on the Internet. We manually removed news items that were not published in Sri Lanka (e.g., the key words may have been used, but the news item was published elsewhere and not aimed at an audience in Sri Lanka). The corpus contains 61986 words in total.

IV. Analysis

For this preliminary analysis, we computed the frequency of occurrence of all the words in the corpus and also collocations that go with the words, *Covid-19* and *Coronavirus*. Word categories such as prepositions, articles, and auxiliary verbs were

excluded in the analysis. In the next section, the findings are presented with a discussion of the different types of words that emerged.

V. RESULTS AND DISCUSSION

A. Word Frequencies

Table 1 presents the most common 25 words in the corpus with frequencies. As can be seen, *COVID-19* and *Coronavirus* are the two most common words that occurred in the corpus. It is understandable as those words were used as key words in the search. What is important is the frequencies of other words. It is possible to categorise all these words as health-related terminology (*COVID-19, Coronavirus, case, patient, health, virus, pandemic, outbreak, death*), administration or pandemic management related terminology (*country, government, people, Colombo, quarantine, travel, area, president*) and economic activities related terminology (*tourism, industry, sector, service*). It is possible to assume that *March* and *April* were referred to the months of the first wave of the pandemic in Sri Lanka.

It is unsurprising that health related vocabulary dominates the corpus. As Ezzarrouki (2020) also highlights, *coronavirus* is a very frequent word used in the Sri Lankan context as well. However, apart from *quarantine*, this set of words is not compatible with the frequent vocabulary identified in other contexts. This could be due to the fact that the first wave of the pandemic in Sri Lanka did not see a large number of cases and deaths and as such certain terms such as *self-isolation* and *flattening the curve* were not very relevant during this period.

The administration or pandemic management related vocabulary highlights government involvement in pandemic management. Economy related vocabulary seems to indicate the impact of COVID-19 on various sectors in Sri Lanka including tourism which is one of the prominent income generators of the country and probably the worst hit.

Two words on the list below stand out, which are *curfew* and *police*. In none of the previous studies, these words have been associated with COVID-19 vocabulary. This is an example of localization of COVID-19 terminology because *lockdown* was termed as *quarantine curfew* in Sri Lanka, and it was imposed by the *Police*.

Table 1: Common words with frequencies

Word	Frequency
COVID-19	495
case	273
Coronavirus	270
country	263
government	200
patient	182
people	173
March	170
tourism	165
health	143
virus	140
Colombo	138
pandemic	134
Police	129
day	124
curfew	118
quarantine	116
April	109
outbreak	107
travel	98
industry	93
area	91
death	91
president	84
sector	83

B. Collocations

As highlighted in our analysis above as well as in other contexts, *Coronavirus* and *COVID-19* are frequently used words in the current English lexicon. It seems several collocations based on these two words have also emerged. For example, Ezzarrouki (2020) identifies *Coronavirus outbreak, novel Coronavirus, spread of Coronavirus* and *fight the Coronavirus* as some collocations of *Coronavirus*.

We also analysed the collocations of *Coronavirus* in the corpus and identified four frequent categories: (1) the word *Coronavirus* was modified by adjectives

and nouns, (2) other nouns were modified by the word *Coronavirus*, (3) verbs were used with *Coronavirus* as object and (4) as subject. Table 2 summarizes the findings.

Table 2. Collocations of *Coronavirus*

Category	Collocations
<i>Coronavirus</i> is modified by adjectives and nouns	<i>novel coronavirus, wide-spreading coronavirus, fast-spreading coronavirus, acute respiratory syndrome coronavirus, deadly coronavirus, human coronavirus, Sri Lanka coronavirus</i>
other nouns are modified by <i>Coronavirus</i>	<i>the coronavirus pandemic, the coronavirus outbreak, coronavirus cases, coronavirus fears, coronavirus patients, coronavirus infections, coronavirus disease, coronavirus statistics, coronavirus emergency, coronavirus pandemic coverage, coronavirus testing kits, coronavirus vaccine, coronavirus cases surge, coronavirus symptoms, coronavirus bodies, coronavirus threat, coronavirus victims, coronavirus death toll, coronavirus test</i>
verbs are used with <i>Coronavirus</i> as object	<i>overburden the coronavirus discovered coronavirus, combat coronavirus, contain the coronavirus, coronavirus is destroyed, labelled the coronavirus, fight the coronavirus, tackle the coronavirus, treat coronavirus</i>
verbs are used with <i>Coronavirus</i> as subject	<i>coronavirus hit, coronavirus attended, coronavirus spread, caused by coronavirus, affected by coronavirus, coronavirus including, coronavirus was..., coronavirus has...</i>

Collocations of *COVID-19* in the corpus can be categorized into three groups mentioned above, except the first, that is, *COVID-19* is modified by

adjectives and nouns. We did not find any examples of this category in the corpus. The *COVID-19* collocations are presented in Table 3.

Table 3. Collocations of *COVID-19*

Category	Collocations
other nouns modified by <i>COVID-19</i>	<i>COVID-19 cases, COVID-19 pandemic, COVID-19 death, COVID-19 outbreak, COVID-19 virus, COVID-19 health care, COVID-19 victims, COVID-19 clusters, COVID-19 impact, COVID-19 complications, COVID-19 test results, COVID-19 crisis, COVID-19 risk, COVID-19 Emergency Fund, COVID-19 infections, COVID-19 test</i>
verbs used with <i>COVID-19</i> as object	<i>tested positive COVID-19, confirmed COVID-19, the virus that causes COVID-19, catch COVID-19, combat COVID-19</i>
verbs used with <i>COVID-19</i> as subject	<i>COVID-19 spreading, COVID-19 coughs, COVID-19 is...</i>

The collocation analysis highlights that *Coronavirus* is used as a noun and a noun modifier in the corpus; however, *COVID-19* is mostly used as a noun modifier. More analysis and inclusion of more sources to the corpus may yield different results; therefore, this result should be interpreted with caution.

VI. CONCLUSION

This preliminary analysis of the corpus of electronic news reports published in English in Sri Lanka on the Internet highlights that some common internationally used *COVID-19* related vocabulary is frequently used in Sri Lanka as well. In addition, some localized vocabulary can also be seen emerging. The terms *Coronavirus* and *COVID-19* collocate with other words: *Coronavirus* is used as a noun and a noun modifier while *COVID-19* is mostly used as a noun modifier. Since this is a preliminary analysis, we recommend more in-depth analysis based on a larger corpus to confirm the findings.

REFERENCES

Aarts, B., Close, J., Leech, G., & Wallis, S. (2013). *The Verb Phrase in English (Studies in English language)*. Cambridge: Cambridge University Press.

Bowker, L. (2020). French-language COVID-19 terminology: International or localized? *The Journal of Internationalization and Localization*, 7(1-2), 1-27. <https://doi.org/10.1075/jial.20014.bow>

Ezzarrouki, I. (2020). Linguistic Coinage during COVID-19 Pandemic: Health Care Terminology. *Journal of English Language Teaching and Applied Linguistics*, 2(1), 54-64.

Fan, L., Yu, H., & Yin, Z. (2020). Stigmatization in social media: Documenting and analyzing hate speech for COVID-19 on Twitter. *Proc Assoc Inf Sci Technol.* 57(e313), 1-11. <https://doi.org/10.1002/pr2.313>

Karami, A. & Anderson, M. (2020). Social media and COVID-19: Characterizing anti-quarantine comments on Twitter. *Proc Assoc Inf Sci Technol.* 57(e349), 1-4. <https://doi.org/10.1002/pr2.349>

Oxford English Dictionary. (2020). [OED.com/view/Entry/88575495](https://www.oed.com/view/Entry/88575495)

Upadhyay M.K., & Maroof K.A. (2020). Understanding the emerging and reemerging terminologies amid the COVID-19 pandemic. *J*

Family Med Prim Care, 9, 5881-7. <https://www.jfmpc.com/text.asp?2020/9/12/5881/305631>

Valentine, S., Mercier, A., Lancelot, R., Roche, M. & Arsevska, E. (2020). Monitoring online media reports for early detection of unknown diseases: Insight from a retrospective study of COVID-19 emergence. *Transbound Emerging Diseases*. DOI: 10.1111/tbed.13738

Xue J., Chen J., Chen C., Zheng C., Li S., Zhu T. (2020). Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. *PLoS ONE* 15(9): e0239441. <https://doi.org/10.1371/journal.pone.0239441>

AUTHOR BIOGRAPHIES



Kithsiri Amaratunga is a Senior Lecturer Grade I, at General Sir John Kotelawala Defence University, and he is the current Dean of the Faculty of Management, Social Sciences and Humanities.



Major Wimansha Abeyawickrama is currently employed at General Sir John Kotelawala Defence university as a Senior Lecturer Grade I, and he is the current Head of the Department of Languages.