

## A Specialized Corpus Study of Lexeme Formation in English Language

Amna Ejaz

MS English Linguistics, International Islamic University, Islamabad.

[amnaejazjutt755@gmail.com](mailto:amnaejazjutt755@gmail.com)

Nighat Shakoor

Assistant Professor, International Islamic University, Islamabad.

[nighat.shakur@iiu.edu.pk](mailto:nighat.shakur@iiu.edu.pk)

### Abstract

Lexicology is a linguistic science that focuses on the theories of morphology and lexeme formations and linguistic components such as lexemes, phrases, and terminologies. *COVID-19* emerged as a global catastrophe encompassing every sphere of life and has had a profound effect on language by giving birth to hundreds of new terms. The present study aims to investigate the lexemes used in *the DAWN* newspaper during *the Covid-19* pandemic. Lieber's model of morphological analysis has been adopted to reach the conclusions. Text for morphological analysis is selected from *the Coronavirus* section of *Dawn* e-newspaper carrying the information related to *the Covid* epidemic, its spread, prevention, and cure, during the time tenure of Sep 2019 to January 2022, hence making it two years long corpus of *COVID-19* data covering first, second, third and fourth waves of the epidemic. The research corpus was further converted into Machine readable format to process it further in *Antconc 3.5.8* software. A total of 82 new lexemes were collected manually from 106 selected articles whose frequencies were studied in *Antconc 3.5.8*. Mixed method research was used to achieve the quantitative and qualitative results. Results demonstrate that four prominent main methods, affixation, compounding, blending, and initials/ acronyms are frequently involved in developing lexical linkages and morphological unities in the coinage of coronavirus lexemes during *COVID-19*.

**Keywords:** Lexicology, Morphology, Lexeme Formation, *COVID-19*, Corpus Analysis

### Introduction

Language evolves from the reasonable external reality which creates it (Halliday, 2007). The year 2020 has been the year of a global pandemic because a lethal and previously unknown virus has caused chaos across the globe by killing 1.5 million people (Dawn, 2020). In the difficult time, the World Health Organisation (WHO) is essential to disease management, prevention, and advice. It makes COVID-19 information available in six American languages that include; French, Spanish, Urdu, Arabic, Chinese, and Russian, and three additional languages such as German, Hindi and Portuguese, to bridge the information gap (Piller et al., 2020).

The COVID-19 virus's influence on the media is inconceivable to disagree. The word coronavirus first appeared in 2019, when WHO first coined its name. Numerous newspapers utilized the latest lexemes in the scenario of a pandemic but the current study intends to investigate the lexeme formation through the COVID-19 pandemic and the emergence of new lexemes in the language of newspapers. Lexicology scrutinizes the chronic patterns of semantic, phonological and morphological or contextual relations. Lexicography is the discipline of lexicology in which the investigation of words is the common subject and it involves word-forming morphemes and word-groups or phrases (Ginzburg, 1966). Ferdinand de Saussure identified two relevant fields of lexicology, general lexicology and specific lexicology, both concerned with the general study of words and vocabulary.

In the book *Introducing Morphology*, Lieber (2009) writes that Americans adopt vocabulary from diverse new concepts such as borrowing, coinage, clipping, blending, functional shift, affixation, and semantic change. In America, lexis study from periodicals and magazines has been a rich source for studying lexis and vocabulary. American lexicologists collected the data from periodicals and magazines and further developed the list of lexis alphabetically. After the collection of the lexemes, they developed the data and then went for standardization (McMillan, 1942). According to the *Longman Dictionary of Applied Linguistics*, word formation is the advent of new words and refers to a morphological process (Richards et al., 1986). Yule describes word formation as a technique of forming and creating new words from already existing words (Yule, 2009).

Lieber's framework of morphological analysis is adopted to analyze the lexeme formation in the DAWN newspaper. This research emphasizes the specialized corpus method for descriptive research in the text of the local newspaper Dawn to investigate the lexeme formation which recently emerged during a global pandemic in 2020. Corpus is the compilation

of a large data that includes naturally occurring text, chosen from a variety of languages (Sinclair, 1991). Similarly, Flowerdew (2012) points out the distinctive features of corpus data- first, it is an empirical methodology and second, it has a phraseological approach to language.

### **Problem Statement**

The present research intends to explore and analyse the process of the formation of new lexemes during the epidemic COVID -19 and the effect of the coronavirus on the development of new lexemes in the English Language, using Lieber's (2009) model of morphology under descriptive Lexicological approach.

### **Research Questions**

- 1) How have the new lexemes been formed in the English language during the outbreak of COVID-19 according to a morphological approach of lexicology?
- 2) To what extent has the English Language in *Dawn* newspaper revealed certain significant lexicological changes in the formation of new lexemes in COVID-19 ?
- 3) How does *Lieber's model of morphology* help analyse socio cum linguistic phenomenon through the English Language used in *Dawn* newspaper during COVID-19 ?

### **Literature Review**

Lexicology is a discipline of linguistic science that is a product of two Greek morphemes, *lexes*, *word* or *phrase* and *logos - learning*. Lexicology has diverse methods to carry out linguistic research that emphasise the use of vocabulary concerning its origin, development and use. Lexicographers theoretically distinguish between *special lexicology* and *general lexicology* (Yule, 2009).

*Lexicalization* is a theory under the realm of lexicology in which lexical innovations and linguistic hierarchies are explored. The theory of Lieber encompassed *lexeme-formation processes* such as affixation, compounding, or borrowing, which are a necessary precondition for lexicalization (Lieber, 2004).

After World War II, COVID-19 was the next catastrophe that impacted every domain of life such as social, political, cultural, and psychological domains. Every century has marked pandemics such as malaria, tuberculosis, influenza, and smallpox but in the realm of pandemics. COVID-19 was drastic among all because it spread far more quickly than previous pandemics. The development of COVID-19 also caused the development of language. WHO

officially confirmed this COVID-19 disease as a pandemic in March 2020 since it spread across 114 countries in just four months and affected around 2 million people.

Mediums of disseminating news were utilizing new linguistic expressions such as lexemes, and lexical unities/phrases. As the phenomenon is recent so with the advancement of disease and its spread language is constantly emerging. Electronic media, print media, and the speech of authorities played their part in governing and utilizing COVID-19-specific terminologies. In keeping the view of drastic social change, newspapers were actively participating to help people understand the situation and its seriousness. The researcher purposefully selected the language of the local Dawn newspaper because of its strong public demand.

In previous research, Covid-specific lexicology has investigated in Arabic and Nigerian languages, whose language sample was the everyday use of language or the actual comments under social media posts. To bridge the information gap in the field of lexicology, researchers looked into the Covid lexemes in local Pakistani newspapers. From a theoretical perspective, the researchers contributed to lexeme formation and lexicology to give an overview of the theoretical foundations underpinning research. The researchers tried to elaborate on *Liebers'* theory of morphological analysis by adding previously established research on the model to measure its reliability.

Hale and Keyser (1993) argued about the structure of lexemes, syntactic structures and their lexical entries but later Lieber (2004) discussed syntactically derived words which was a paradigm shift in the domain of lexicology and lexicography. Lieber (2009) presents lexicology as a part of grammar rather than presenting it as a separate component. Lexical entities provided by Lieber indicate category variation, phonological information, and semantic information. His model appreciates the diverse level of productivity of lexemes rather than following a pre-described way. Unlike previous methods and approaches to morphology, Lieber's model provides an in-depth analysis of lexemes and lexical processes which includes morphology too. Research adheres to Lieber's model of morphological analysis in a way that follows a systematic process to explain the creation of lexemes which can be applied to every language equally. His model encompasses French, Dutch, and English as well. Morphological rules can be applied to lexemes just like phrasal structures are discussed under syntax (Sugioka, 1992).

Socio-historical and socio-cultural events drastically impacted language and results can be seen in the text of the newspaper, articles and research reports hence making the changes observable in the linguistic body of the newspaper. The researchers studied and examined the lexical shift which occurs in the *DAWN* newspaper particularly the linguistic changes due to the emerging linguistic trends during *Coronavirus*. The study sought to build up a rationale to solidify the changes in language and the way these changes have affected the language at large. Thus, due to more reliance on technology, the present research has utilized the electronic corpora for conducting its analysis.

Linguists, lexicographers, sociolinguists, and researchers are keenly interested in the systematic change that Covid-19 has brought in language. A professor at King's College London, Tony Thorne (2020) wrote an article under the heading '*The language of COVID-19 goes viral*' and states that everyday language was also transformed by the coronavirus pandemic using the hashtag *Corona speaks* (Thorne, 2020). In constant confusion and social changes during COVID-19, lexicographers have witnessed a drastic explosion of new terminologies, lexemes, and phrases (Katermina & Yachenko, 2020).

Socio-linguistic change of COVID-19 brought a clear change in the Nigerian language which emerged during this time due to the impact of COVID-19. The research included two thousand Nigerian speakers in Southwest Nigeria and results showed different uses of morphological processes in the creation of neologism in Nigerian English (Kupolati et al., 2021).

Bloomfieldian sense lexicon as a repository, by considering morphological lexicon, is proposed as the set of potential lexemes of a language. Morphological results shed light on explaining the syntactic patterns within sentences by morphological approach. Research on Spanish newspapers was conducted to build synchronic dictionaries based on the corpora of the living lexicon. The research was focusing on the birth of words: neology and removal; the death of words; and desuetude. Chomsky's lexical hypothesis states that words are collected both in the mental and collective lexicon of an individual speaker or a linguistic community (Aronoff, 2018).

A lexicon is a special filter through which the words have to be passed after they have been generated by the word formation rules. He further suggested that the list of morphemes together with the rules of word-formation define the set of potential words of the language.

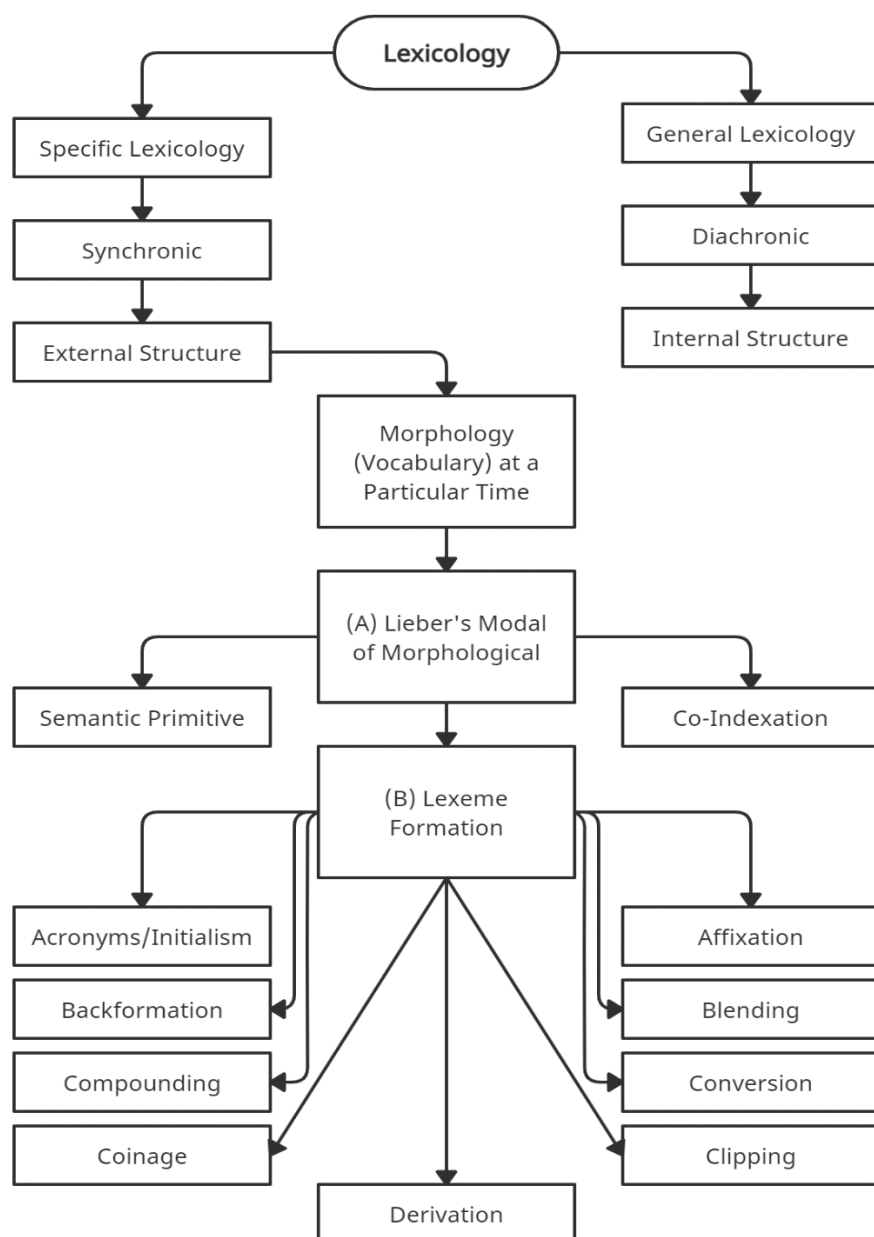
By using a statistical method, Bonami and Thuilier (2019) describe duplicate rivalry in lexeme creation as a situation in which several lexeme formation procedures are employed to fill in morphological unity gaps. Their research focuses on the rivalry between two suffixes -*iser* and -*ifier* in French which derive verbs from two grammatical categories: nouns or adjectives. Affixation is one of the most prominent forms of the lexeme formation process and it is used with nouns, adjectives and verbs. Affixation can have -action, -ment, -al, -ity, -ness, -less, re, -ee etc. Language usually has affixation which falls under specific categories.

- *Personal affixes* create 'people nouns' from verbs or nouns. In this category, affixes in English are the suffix -er and the suffix -ee in the words writer and employee.
- *Negative and privative affixes* are the affixation types that add a meaning of negation i.e. not to their base. Examples in English include, un-, in-, and non-.
- *Prepositional and relational affixes* express notions of space and time. Examples include prefixes like *over-* and *out-*.
- *Quantitative affixes* are the type of affixes which express words about amount and quantity like -ful and multi- in the lexemes such as helpful, and multifaceted.
- *Evaluative affixes* express a smaller version of the base such as -let as in couplet or leaflet. Unlike evaluative affixes, *Augmentative* affixes have bigger base versions like mega- (megastore, mega bite). Diminutives often convey affection or endearment.

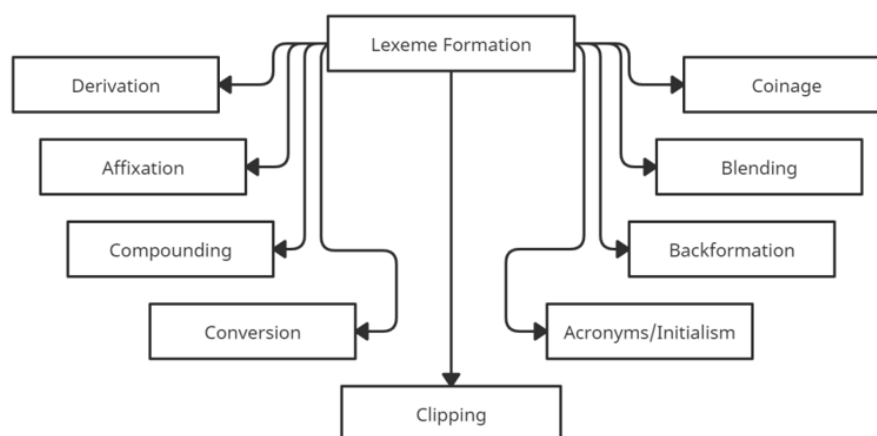
Lieber's *Introducing Morphology*(2009) has discussed three types of compounding prominent in lexeme formation and morphology. One common way of dividing the compound categories is into *root compounds* and *synthetic compounds*. Lexemes according to their semantic and grammatical categories, include three prominent types of relation: *attributive compound*, *coordinative compound*, and *subordinative compounds* (Lieber, 2009).

Lieber's theory (2009) deals with affixes that are lexical items containing both phonological and morpho-syntactic features. The researcher agrees with Lieber that both stem and affixes are lexical entities connecting morpho-syntactic and phonological features. They draw a comparison between DM syntactic operations and Lieber's lexical morphology. First, DM syntactic operations combine terminal roots to create words. Secondly, in DM theory, morpho-syntactic features are involved in syntax created by vocabulary insertion. DM agreed on a major insight into Andrews' theory and compared it with Lieber's theory.

The given self-devised Figure 1 demonstrates the elements of lexicology & Lieber's model.

**Figure 1***Research model and framework***Lieber's Theoretical Model of Lexicology**

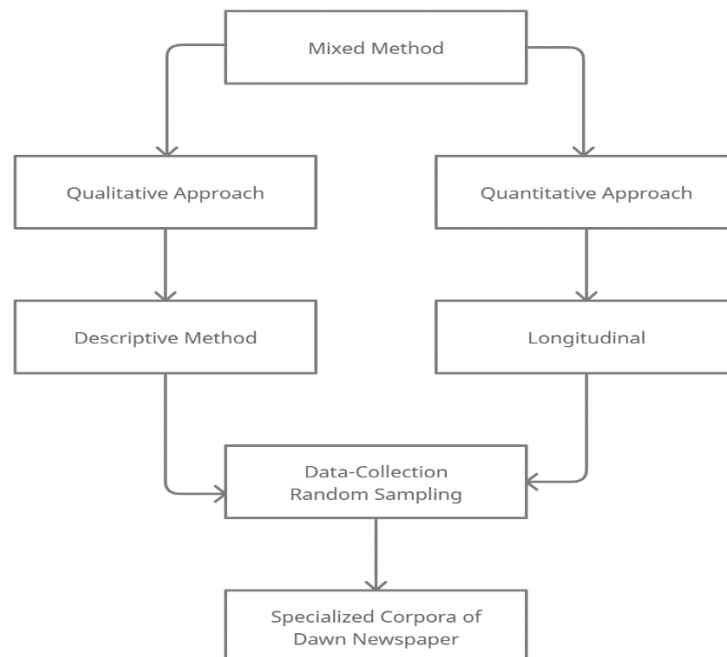
Lieber has underlined the scheme of lexeme formation. The nine distinct types of word formation and sub-types by Lieber (2009) can be presented diagrammatically in Figure 2:

**Figure 2***Lexeme formation processes*

### Theoretical Framework and Methodology

Mixed-method has been adopted as a methodology for the analysis which constitutes both quantitative and qualitative methods. In the quantitative method, the concern of the study is to adhere to a corpus-based methodology whereas, in qualitative method research, the researcher intends to analyze descriptive patterns and methods presented by Lieber. The mixed method provides the framework for the analysis of lexemes via lexeme formation processes in the scenario of a local *DAWN* newspaper during the COVID-19 pandemic. Lieber's *Lexeme-based Morphology theory* has been adopted for the investigation of neologism presented in the collected corpus of *DAWN*. As the research is corpus-based and is termed a specialized corpus by Kuebler and Zinsmeister (2015), so data for the research is gathered through random sampling. Antconc 3.5.8 version of the software has been selected as a corpus tool to mark the frequencies of novel lexemes. Language data is taken from the period of COVID-19 from Jan 2021 to Jan 2022. The corpus collected is termed a *specialized corpus* which specifically consists of Dawn newspaper containing about 106 from *EOS*, Coronavirus, and editorial sections of the newspaper covering the domains of economy, global politics, medicine, education, domestic life, and so on. The methodological framework of the research drawn from Lieber's model of Morphology is presented diagrammatically in Figure 3:



**Figure 3***Diagram of Liber's model of Morphology***Data Collection: Dawn e-Newspaper**

- As Covid was an emerging issue, books were not published which could define the vocabulary and lexicons of the current pandemic. A written record of the issue was not present. In that scenario, newspapers were the prime source of spreading Covid news, and specific language was used, particularly vocabulary and lexicons.
- About 21 English newspapers are published in Pakistan. Amongst them, some are published daily while others are published weekly, and no one among them was providing comprehensive coverage of the theme of COVID-19 as compared to the *Dawn* newspaper.
- *Dawn* newspaper is the oldest and is considered among the largest newspapers in the country whose credibility and authenticity have yet not been questioned.
- *Dawn* newspaper provides the most comprehensive and extensive coverage of the theme of COVID-19 in the 'coronavirus section' as compared to the other local English newspapers.

## Corpus Collection

Covid data from *e-Dawn* was collected from September 2019 to January 2022 for almost 02 years. The timespan for data collection is carefully considered because it covers 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>, and 4<sup>th</sup> waves of COVID-19 across the globe. About 106 Articles were downloaded from *e-Dawn* containing COVID-19 information which was further converted into a machine-readable notepad file through *Antconc 3.5.8* software that can process only notepad files. Among 106 articles covering coronavirus information, 82 novel words/lexical unities emerged into language.

## Analysis

Mixed method research adheres to quantitative and qualitative research under one framework. Quantitative analysis is carried out by utilizing *Antconc 3.5.8* software, particularly for the frequency of novel lexicons that represent the text. The qualitative section of the research considers the descriptive method title, *Lieber's Model of Morphological Analysis* which comprises of following processes: derivation, conversion, affixation, acronyms/initials, compounding, coinage, clipping, blending, and back formation.

COVID-19 is a combination of 'Coronavirus disease' and 19 is the year 2019 which marked the initiation and emergence of deadly virus. On 11<sup>th</sup> February 2020, COVID-19 terminology was coined by WHO. Vocabulary terms containing 'covid' is a lexeme derived from words such as *coronacation*, *covexit*, *coronacoma*, and *covidiot*.

## Defining the Research Corpus

In Corpus Linguistics, data is electronically available texts of language related to any linguistic realm or language variety. 'Corpus' is synonymous with electronically available language text (Kuebler & Zinsmeister, 2015).

## Concordances

Antconc software contains lines when we search for a word to help search words. *Key Word in Context (KWIC)* is for concordance. Searched words appeared on the Antconc screen with respective frequencies and that is visible in Figure 4. A total of 293 concordance hits are displayed from the lexeme.

**Figure 4***Antconc 3.5.8 Software and concordance hits of COVID-19*

of the total 1,450 hospitalised people due to Covid-19. A health department report said that of  
 , it is possible to get infected with Covid-19 a few days before or after getting  
 that vaccines could not infect people with Covid-19, adding that 'inactivated vaccines don't  
 by the NCOC in the fight against Covid-19, adding that it was a classic example  
 that it is not possible to contract Covid-19 after receiving the Sinopharm vaccine as it  
 pay for failing to get vaccinated against Covid-19 ahead of a deadline this week, Health  
 largest women's college have contracted Covid-19, an official from the Federal Directorate of  
 care infrastructure is frighteningly real. Read: Covid-19 and thereafter In India, the second wave  
 the technology used to produce and formulate Covid-19 and other mRNA vaccines. Joe Biden has  
 cent of those vaccinated and prevented severe Covid-19 and hospitalisation in 100pc of the studied  
 government's Scientific Task Force on Covid-19 and Vice Chancellor University of Health Sciences  
 , which has recorded more than 559,000 cases of Covid-19 and more than 12,100 deaths, is still largely  
 private school in H-8 - tested positive for Covid-19 and are in isolation. The three institutions  
 to crack down on health misinformation around Covid-19 and other diseases. Video-sharing giant YouTube  
 ity, severity, and other features associated with COVID-19 and investigations are ongoing." 'Reckless to as  
 30 per cent of the targeted population, against Covid-19 and will take the target to 40 per  
 federal government's Scientific Task Force on Covid-19 and Vice Chancellor University of Health Sciences  
 straining to contain a second wave of Covid-19 are turning to faster, cheaper but less  
 operating procedure to stem the spread of Covid-19 as the fourth wave of the virus  
 come back' Experts warn that even if COVID-19 becomes less active in the summer, it  
 reported in Nowshera, while the occupancy of Covid-19 beds was 6.1 per cent. It added that 46  
 reported in Nowshera, while the occupancy of Covid-19 beds was 6.1 per cent. It added that 46  
 an easy-to-administer antiviral pill for Covid-19, but so far only antibody cocktails which  
 to synthesise mRNA vaccines, not just for Covid-19 but for many other widespread infections like

**List of Lexemes- Frequency Count**

A total of 82 novel lexical terms are collected through careful reading of 106 articles randomly chosen from Dawn newspaper. Articles are further listed in notepad files and tables present their number of occurrences. In Antconc, 'word lists' allow for examination of the text in concordance hits and the frequencies that appear there.

**Labelling and Categorization of Lexical Elements**

Frequencies in the data present the co-happening linguistic features of lexemes. A diversity of new terminologies emerged daily, which illustrated the linguistic significance of the COVID-19 pandemic. In addition, new lexical entrances have been perceived during the endemic and few longstanding lexemes were adopted in a new context. In the 17<sup>th</sup>, 18<sup>th</sup>, and 19<sup>th</sup> centuries, some words were used but they got adopted and frequently used in recent pandemics and had no longer the same meanings.

Due to the progressions in COVID-19 vaccination, protections, antibiotics, and healthcare, numerous developments were made daily which ended up having new

terminologies. Print and electronic media used these terminologies globally. The present research thus endeavours to highlight the new formations of lexemes during the pandemic and the process and the impact the coronavirus had on language development. In this regard, special emphasis has been given to the language used in *the Dawn* newspaper. The given table 1 provides a gist of such English lexemes related to health and medical aspects:

**Table 1**

*Medical/Health lexemes from e-Dawn*

Medical /Health lexemes from e-Dawn			
Moderna (Vaccine)	SARS-CoV-2 (vaccine)	Remdesivir (antiviral)	Adino-vector (injection)
Immunocompromised	Merck Pill	Pfizer-BioNTech	Sinopharm (vaccine)
AstraZeneca	Cansino (vaccine)	Alpha Variant	SinoVac (vaccine)
N-95	Immunogenic (Antigen)	Post-Vax	Molnupiravir (Antiviral Medication)
KN95	Anti-vax	Veklury (Antiviral Medication)	Booster Jab
CoronaVac (vaccine)	Delta-infected	Delta variant	Delta Plus
Omicron variant	Lamba variant	Omicron (Covid Variant)	Monoclonal
Medicines Patent Pool (MPP)	PPE (Personal Protection equipment)	Vaxzevria (vaccine)	Sputnik V (Covid vaccine)
PCR (Polymerase Chain Reaction)	Delta Strain	NGS (Next Generation Sequencing)	Coronavirus
COVID-19			

Terminologies related to socio-economic scenarios are related to protection, prevention, and collective attitude during the outbreak and are presented in the given table 2.

**Table 2***Social, Educational, and Economic Terms form the research corpus*

<b>Social, Educational, and Economic Terms form the research corpus</b>		
Covidiot	SOPs (Standard Operating Procedure)	Covid era economy
Social Distancing	Anti-coronavirus	Incubation centers
(RED) Reach Every Door	New Crude normal	Cyberconsultation
Anti-vaxxer	Post-Covid	Weekend Curfew
Coronation	Pre-pandemic	WFH (Working from home)
SRO (Standing Room Only)	Virus-idled	Covid-era economy
Anti-epidemic	TTQ (Testing, tracing and Quarantine)	Infection Hotspots
Micro Smart Lockdown	Frontline health workers	Anti-Covid
Booster shot	Anti-epidemic	Cordon off
Covid Jab centers	TeleHealth	eDoctor
Infodemic	Telemedicine	TeleDoc

Terminologies like telemedicine, TeleDoc, eDoctor, Cyber consultation, and TeleHealth are the latest emergence of words in the health sector and show a connection between digitalization and health that happened during COVID-19. Due to COVID-19 prevention and lockdown restrictions, visiting hospitals was not safe so doctors were providing online/cyber health facilities. However, ‘infodemic’ is related to education and information. *WHO* explains the lexeme as recursive information that is present digitally but it’s difficult to distinguish between real and fake news. *Merriam-Webster dictionary* added ‘Cordon off’ as a novel lexeme into the dictionary on 2<sup>nd</sup> January 2022. Table 3 presents lexemes that emerged in political/global spheres.

**Table 3***Political/Global Lexemes*

<b>Political/Global Lexemes</b>	
ECDC (European Centre for Disease prevention and control)	NPP (Non-personal Promotion)
EUAs (Emergency Use Authorizations)	TRIPS wavier (WHO International Agreement)
NPIs (Non-Pharmaceutical Interventions)	CDC (Center for Disease Control and Prevention)
NCOC/NCC (National Command and Operation Center)	SAPM
	GAVI

Keywords in Tables 1, 2, and 3 presented the dominant coronavirus lexemes that are the combination of nine (9) different types of word formation processes hence creating new word unities that have meanings in the context of coronavirus.

### Lieber's Model: A Descriptive Method

In *Introducing Morphology* (2009), Lieber mentioned word formation processes amongst the most prevalent processes are affixation, compounding, conversion, coinage, and blending.

Special emphasis is given to processes like affixation, compounding, Initials/acronyms, and conversion as components of Lieber's model. *Affixation* is pertinent when it comes to analysing lexemes based on their constructions. Some pertinent kinds of affixation are *Personal affixes*; *Negative and privative*, *Prepositional*, *relational*, *Quantitative* and *Evaluative affixes*, and *Argumentative affixes*.

Cumulatively, five prominent types of affixation are mentioned by Lieber in the affixations section and lexemes are mentioned based on the sub-types in table 4, whereas table 5 illustrates lexemes according to the respective categories of affixation.

**Table 4**

#### *Categories of affixation*

Categories of Affixation		
	Lexemes	Frequencies in corpus
1	Anti-coronavirus	04
2	Post-covid	01
3	Anti-vex	01
4	Anti-vaxxer	01
5	Anti-covid	02
6	eDoctor	03
7	Pre-pandemic	02
8	Sub-lineage	05
9	Epicentre	01
10	Norovirus	01
	Total	21

**Table 5***Affixations and their types identified in the corpus*

	Negative & Primitive Affixes	Prepositional & relational affixes	Evaluative & Argumentative Affixes
1	Anti-coronavirus	Pre-pandemic	Sub-lineage
2	Anti-covid	Post-covid	Norovirus
3	Anti-vaxxer	Epicentre	Epicentre
4	Anti-vex		

*Note. The model covers Affixation at 5 distinct levels which have been dealt in the analysis part of the study too. Amongst category divisions, few are present in the research corpus and hence are projected and illustrated in the tables.*

**Compounding**

The theory of lexicology discussed by Lieber in *Introducing Morphology* provides three kinds of compounding noticeable in morphology specifically in lexeme formation. *The attributive compound* is the first one that takes nouns as modifiers of the head. i.e windmill. *The coordinative compound* is the second type in Lieber's lexeme formation but lexemes/words have equal weightage here likewise: doctor-patient, and producer-organiser. Third is a *subordinative compound* which is derived from a verb or an element of a verb like; -er, -ing, -tion, -ment, etc. (Lieber, 2009). Following is the list of compounding in tabulated form:

**Table 6***Frequencies of Compounding*

Frequencies of Compounding		
	Compounds	Frequencies
1	Frontline health workers	02
2	Delta-infected	01
3	lambda variant	01
4	Delta variant	40
5	Omicron variant	09
6	Covid Jab centres	01
7	Weekend Curfew	01
8	Delta Plus	12
9	Booster Jab	11
10	Cordon off	01
11	Smart lockdown	08
12	Infection Hotspots	02
13	Covax Scheme	04
14	Reach every door (RED)	02
15	New crude normal	01
16	COVID-19	293
17	Covid positive	03

18	Alpha variant	02
19	Merck pill	02
20	Booster jab	11
21	CoronaVac	08
22	Adino-vector	03
23	Mass testing	04
24	Health Pass	03
25	AstraZeneca	25
26	Homeschooling	14
27	Social distancing	18
28	Booster shot	09
29	Herd immunity	17
30	Super Spreader	05
31	Community spread	01
32	Cyber consultation	01
33	Immunocompromised	04
34	Coronavirus	175
35	Vaccination Apartheid	02
	Total	698

**Table 7***Types of Compounding*

Types of Compounding			
	Attributive	Coordinative	Subordinative
1	Omicron variant	COVID-19	Self-isolating
2	COVID-19 positive	Coronavirus	Mass testing
3	Delta infected	Lamda variant	Infection hotspots
4	Frontline health workers	Delta Plus	Social distancing
5	Covax scheme	Smart lockdown	Incubation centres
6	Super spreader	Covid Jab centres	Home Schooling
7	Booster shoot	Weekend curfew	Immunocompromised
8	Herd immunity	Booster Jab	Cyber consultation
9	Merck pill	Alpha variant	Incubation period
10	CoronaVac	Loose lockdown	
11		New crude normal	
12		Reach every door	
13		Adino-vector	
14		Health pass	



15	Beta variant
16	Corden off
17	Vaccination Apartheid

## Blending

The predominant word formation process specifically in the corpus of *Dawn* is blending. SinoVac, covidiot, coronacation, infodemic, and immunogenic are some terms that emerged with the process of blending. *SinoVax* is a blend of ‘Sino’ and ‘Vac’. Similarly, *Covidiot* is a combination of ‘covid and idiot’- a person who is not following *SOPs* and is considered an idiot. Whereas, *infodemic* is from ‘information and pandemic/epidemic’. *Immunogenic* is combined with ‘immunity’ and ‘genes’. Lastly, *immunocompromised* combines ‘immune + compromised’. Lexemes emerged from blending and their frequencies are tabulated in Table 8:

**Table 8**

### *Blending and its frequencies*

Blending		
	Lexemes	Frequencies
1	SinoVac	34
2	Covidiot	01
3	Infodemic	04
4	Coronation	02
5	Cansino	17
6	CoronaVac	08
7	Pfizer-BioNTech	04
8	Remedesivir	05
10	Molnupiravir	20
11	Moderna	16
12	Epidemiology	04
	Total	115

## Clipping

It is a category of lexeme formation that is meant to shorten lengthy lexemes into single syllables. *CoronaVac*, *Tele-health*, *Cansino*, *Sinopharm*, and *KN95* are the examples. Lexical constructions via Clipping are shown in the Table 9:

**Table 9***Clipping and its frequencies*

<b>Clipping</b>		
	<b>Lexemes</b>	<b>Frequencies</b>
1	SinoPharm	35
2	N95	16
3	KN95	01
4	Tele-doc	01
5	Tele-Health	01
6	Tele-medicine	01
7	Tele-consultation	01
	Total	56

**Coinage**

It is a process of establishing new lexemes by combining completely dissimilar word mixtures that do not have semantic relevance. Coinage and its number of appearances are presented in the given table:

**Table 10***List of Coinage*

<b>List of Coinage</b>		
	<b>Lexemes</b>	<b>Frequencies</b>
1	Sputnik V (Covid vaccine)	05
2	Veklury	01
3	Vaxzevria (vaccine)	06
	Total	12

**Acronyms/Initials**

Initials/acronyms are dominantly used to give names to larger word groups and organisations. In the context of COVID-19, these initials were used to name the institutions and organizations that possess the authority. Lexical-Formation from Acronyms/Initials are presented in tabular form below:

**Table 11***Frequencies of Acronyms*

<b>Frequencies of Acronyms</b>		
	<b>Lexemes</b>	<b>Frequencies</b>
1	NCC (National Command and Operation Center)	04
2	TRIPS wavier	01
3	CDC	16
4	SAPM	13
5	GAVI	05
6	ECDC	01
7	KN95	01
8	NCOC	65
9	NPIs	06
10	(RED) Reach Every Door	08
11	SRO (Standing Room Only)	01
12	N95	16
13	SOPs (Standard Operating Procedure)	32
14	TTQ (Testing, tracing and Quarantine)	01
15	NGS (Next Generation Sequencing)	06
16	PCR (Polymerase Chain Reaction)	04
17	PPE (Personal Protection equipment)	05
18	NPP (Non-personal Promotion)	02
19	EUAs (Emergency Use Authorizations)	02
	Total	188

In conclusion, Covid-pandemic times have been illustrated and explained through Lieber's model. Following are the frequently used word groups in *Dawn* during 2 years of covid-pandemic. Table 12 below lists the frequencies of these lexemes.

**Table 12***Cumulative frequencies and lexical constructions are mentioned below*

<b>Lexeme Formation Types by Lieber</b>		
<b>Classifications of Lexeme Formation</b>	<b>Quantity of lexemes</b>	<b>Cumulative Frequencies</b>
1) Derivation	00	00
2) Affixation	07	14
3) Compounding	35	498
4) Conversion	00	00
5) Coinage	03	12
6) Blending	11	111
7) Back Formation	00	00
8) Acronyms/ Initialism	19	186
9) Clipping	07	56
Total	82	879

After examining the frequency count of lexemes in Table 12, the direct application and employment of Lieber's model in lexeme formation have been verified. Results validate the process of lexeme formation as an outcome of COVID-19. The tables above enlisted

cumulative frequencies of the lexeme in different categories of lexeme formation besides the dominant processes.

## Findings

During COVID-19, numerous vocabulary/lexemes were produced in language that had no pre-pandemic existence. This study explicitly examined drastic changes in the linguistic body of the English language by utilizing the model of the prominent theorist and lexicologist 'Lieber' in the field of lexicology and lexicography. In addition, the *DAWN* newspaper is widely read and understood in Pakistan and its language has an impact on the overall scenario of language adaptation. As this is a corpus-based descriptive study, so quantitative method is utilised along with a qualitative method to answer the research questions.

To answer Question 1 of my study, I chose Lieber's model of Morphology which provided a comprehensive framework to analyse morphological and lexical elements in language by giving nine (9) key types of processes of lexeme formation. Research question 1 highlights the importance of COVID-19 in bringing a shift into the language, observed through *DAWN* newspaper articles. The aforementioned nine (9) lexeme formation processes provided a theoretical framework to answer question no 1 and within 106 articles randomly collected from *DAWN* newspaper, the researchers manually listed new words which came up due to novel COVID-19.

Question 2 of this research raised a point about the significance of the emergence of lexemes in language. It intended to investigate the lexical coinage that came up as a result of Covid-19 in the corpus. Lexemes are mostly related to health care, vaccination, social scenarios, and global affairs. Lieber's key tenets have provided points which categorize lexeme further in their sub-parts. The English language witnessed the involvement of certain processes in lexeme formations during the COVID-19 pandemic. Lieber's nine (9) key tenets from his theory of morphology provided theoretical support. Amongst those types, four (4) of them are predominantly used in lexical formations, therefore researchers emphasize these lexeme formations. Words such as social distancing, new crude normal, working from home, homeschooling, covidiot, coronacation, eDectors, telehealth, post-vax, Pfixer, COVID-19, Coronavirus etc. are lexicons that have no pre-pandemic existence. As a result of the COVID outbreak, 82 new lexemes have been explored in the selected corpus of Dawn newspapers.

From the corpus of *Dawn* newspaper articles, I generated a manual list of lexemes and further study was carried out through *Antconc* 3.5.8 version that provides statistical data. The manual lexicon list consists of 82 lexemes in total. For statistics, *Antconc* 3.5.8 provides frequencies of each of those lexemes in the data. Lexeme formations are mentioned according to the respective categories provided by Lieber (see Table 12).

Question no. 3 investigates social cum linguistic elements expressed in the *Dawn* newspaper. Results of the study demonstrate that *Dawn* newspaper expressed social cum linguistic phenomenon specifically during the Covid pandemic. To answer the language change, I collected 106 articles from the *Coronavirus*, *Editorials*, and *EOS* sections of the *Dawn* newspaper. The duration of the corpus was a two-year pandemic time that covered the first, second, third, and fourth waves of COVID-19 which affected the social structure at large. Social changes correspond to all the changes which we have adopted and experienced in our lifestyle during the specific COVID-19 years. Corpus is from Sep 2019 to Jan 2022 depicting social changes that caused linguistic changes too because language is not independent of socioeconomic circumstances. The effect of coronavirus on language is huge and the results of the frequencies of lexemes are the evidence of social change that has been experienced. Furthermore, the frequencies of these 82 lexemes are different in the corpus such as lexemes emerged through compounding are 35, acronyms are 19, blending is 13, affixations are 7, clippings are 7, and coinages are 3. The research corpus does not contain any words formed by the use of back-formation and derivation.

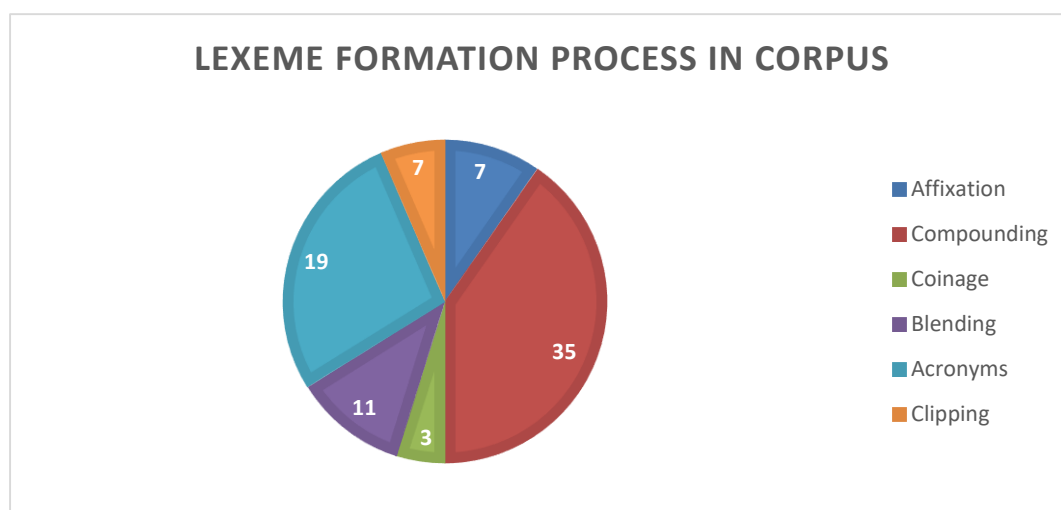
Qualitative research is carried out by categorizing lexemes into the sub-categories of medical, social, political/global, and educational. Lexical additions in these categories are mentioned in Tables 1, 2, and 3. A total of 82 lexemes were collected from the corpus of *Dawn* newspaper and among these 82 lexemes, 37 belong to the medical/health category, 33 belong to the social, economic, and educational domain, and 09 of them belong to the global or political domains. Moreover, lexemes are further divided into the sub-categories of word-formation types; compounding and affixation. I divide the analysis part of the research into 5 sections covering both quantitative and qualitative analysis simultaneously.

If we see the results in Tables 4, 5, 6, 7, 8, 9, 10, and 11 in the data analysis section, we can observe the lists of words that emerged by the process of affixation, compounding, acronyms/initialism, clipping, coinage, and blending. Table 12 mentions the cumulative frequencies of the aforementioned processes and their number of appearances in the corpus.

Results of these lexemes indicate that amongst nine processes of word formation only Compounding, Acronyms/Initials, and Blending dominate the word-formation processes in the present data. Lexemes that emerged due to these processes indicate that Lieber's model of Morphology presents an accurate representation of lexeme formation processes. The result of frequencies also validates that compounding is the most commonly used lexeme formation process in the English language. Some lexemes were already present in the English language but due to emerging changes in the sphere of language during the coronavirus, these previously established lexicons got new meanings by coming together through the process of compounding and now their compounds are carrying meanings in the context of Covid pandemic.

If we rate all the lexeme formation processes, we can express them in a bar graph as follows:

**Figure 6**  
*Outcomes of Lexemes Formation*



In Pi chart 6, lexeme formation processes indicating their respective frequencies of occurrence in the corpus have been presented. Descriptive and statistical analysis of the research indicates that questions raised in this research regarding the emergence of lexemes and Lieber's model have been answered. Lexical entries presented in Tables 1, 2, and 3 are now standardized words and have independent meanings in dictionaries. Likewise, *WHO* has also given meanings to *COVID-19* terminologies such as social distancing, covidiot, coronacation, and so on. Therefore, we can conclude our research on the point that language change involves morphological processes of word formation hence adding new lexemes to the English language.

## Conclusion

The present study is significant in a way that it investigates the most pervasive issue facing the entire globe. It focuses on the most recent lexical formations in the English language. No recent studies regarding lexeme formation have been conducted in the context of local newspapers. It is pertinent since it helps to investigate the morphological processes underlying the lexeme formation and coinage of lexemes in any field during the emergent time of COVID-19. According to Mweri (2021), language change is an occurrence by which alterations are made in the attributes and the use of a language over time. All natural languages can change and these changes are normally reflected in areas of language use. Language changes can be manifested in sound patterns, lexis, semantics, and syntax. Language changes may be systematic or sporadic. During a time of social crises like the COVID-19 pandemic, language of social crisis is used and there is normally an explosion of new words and phrases like those discussed above. The new lexemes are useful since they function to help people communicate their fears about the biggest health crisis ever seen in generations. It collates people around a set of collective cultural reference points – it therefore acts as a kind of lexical "social glue" (Lawson, 2020).

From the discussion, it is also clear that the spread of coronavirus disease changed the lives of billions of people worldwide. Likewise, it ushered a new set of lexicons that encompassed specialist terms from the fields of epidemiology and medicine to the general populace and their social structures. New acronyms were created, and new words were coined to express the societal importance of the imposed isolation and social distancing. The above discussion also indicates that language changes to accommodate new happenings in society—Like the COVID-19 pandemic. The changes discussed above focus on the vocabulary level because grammatical and phonological structures are relatively stable but vocabularies can change very quickly both in words and meanings. A new lexeme formation process has been explored and examined by the researcher, how existing words or expressions have acquired new meaning (semantic shift) or prominence, and how completely new coinages through blending or acronyms have been occasioned by COVID-19. According to Lawson (2020), When there is significant social or civic change during the time of the COVID-19 pandemic, linguistic creativity reflects the major preoccupations of the time and can show how people gather to talk about new challenges and contexts. As long as the coronavirus is on a rampage, understanding the language surrounding it will be ever more important.

Finally, we can say that linguistic change is crucial to investigate in the context of socio-economic factors. Linguists studied linguistic change and they concluded that linguistics should stop considering language as an autonomous body and rather take it as an organism which lives, evolves and dies independently. Accordingly, language depends on the actions of the users of the language and the situation where they use language. So, the research undertaken is a successive addition to the discipline of *Corpus Linguistics* and *Lexicology* because it provides comprehensive data for lexical investigation, the evolution of Lexicology and the way forward for future researchers and scholars.



## References

- AFP. (2020, March 14). Europe is now epicentre of coronavirus, says WHO. *Dawn*.  
<https://www.dawn.com/news/1540774>
- Aronoff, M. (2018). Morphology and words: A memoir. *The lexeme in descriptive and theoretical morphology*, 4(3).
- Bonami, O., & Thuilier, J. (2019). A statistical approach to rivalry in lexeme formation: French-iser and-ifier. *Word structure*, 12(1), 4-41.  
<https://www.eupublishing.com/doi/10.3366/word.2018.0130>
- Ginzburg, R. S., Khidekel, S. S., Knyazeva, G. Y., & Sankin, A. A. (1966). *A course in modern English lexicology*. Higher School Publishing House.
- Flowerdew, L. (2012). Corpora and Language Education. Basingstoke: Palgrave Macmillan.  
*International Journal of Corpus Linguistics* 19(1).  
<https://doi.org/10.1075/ijcl.19.1.06fel>
- Halliday, M. A. K., & Yallop, C. (2007). *Lexicology: A short Introduction*. Continuum. <https://ci.nii.ac.jp/ncid/BA81400100>.
- Hale, K., & Keyser, S. J. (1993). On argument structure and the lexical expression of syntactic relations. *An annotated syntax reader*, 312-327.
- Katermina, V., & Yachenko, E. (2020). Axiology of COVID-19 as a linguistic phenomenon in English mass media discourse. *Advances in Journalism and Communication*, 8(2), 59.  
<https://doi.org/10.4236/ajc.2020.82005>
- Kuebler, S., & Zinsmeister, H. (2015). *Corpus linguistics and linguistically annotated corpora*. Bloomsbury Publishing. <https://doi.org/10.5040/9781472593573>.
- Kupolati, O. O., Adebileje, A., & Adeleke, A. A. (2021). “Someone has been coronated” Nigerian English lexical innovations in the COVID-19 Pandemic. *Cogent Arts & Humanities*, 8(1). <https://doi.org/10.1080/23311983.2021.1947559>
- Lawson-Adams, J. (2020). Building Lexical Representations with Nonverbal Supports. *International Literacy Association*. 56(3). <https://doi.org/10.1002/rrq.326>
- Lieber, R. (2004). *Morphology and lexical semantics*. Cambridge University Press.

Lieber, R. (2009). *Introducing Morphology*. Cambridge University Press.

<https://doi.org/10.1017/cdo9780511808845>.

McMillan, J. (1942). American Lexicology. *Jstor.org*, 53(2), 141-163.

<https://doi.org/10.2307/454926>

Mweri, J. (2021). Corona Virus Disease (COVID-19) Effects on Language Use: An Analysis of Neologism. *Linguistics and Literature Studies*. 9(1). 36-47.

DOI: 10.13189/lls.2021.090105

Richards, J. Platt, J., & Weber, H. (1986). Longman Dictionary of Applied Linguistics. *Interlangauge Studies Bulletin*, 2(1). <https://doi.org/10.1177/026765838600200107>

Sinclair, J. (1991). Corpus, Concordance, Collocation. *Oxford University Press*.

Sugioka, Y. (1992). Deconstructing Morphology. *J-Stage*. 11, 245-266

<https://doi.org/10.9793/elsj1984.11.245>

Piller, I., Zhang, J., & Li, J. (2020). Linguistic diversity in a time of crisis: Language challenges of the COVID-19 pandemic. *Multilingua*, 39(5), 503-515.

Thorne, T. (2020). #CORONASPEAK-the language of Covid-19 goes viral. *Kings College London*. <https://www.kcl.ac.uk/news/coronaspeak-the-language-of-covid-19-goes-viral>

WHO. (2020). Coronavirus disease (Covid-19) Pandemic.

<https://www.who.int/emergencies/diseases/novel-coronavirus-2019>

WHO. (2020). WHO Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020. *World Health Organization*. <https://www.who.int/director-general/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020>

Yule, G. (2009). *The Study of Language*. Cambridge University Press. <https://doi.org/10.1017/9781009233446>.