Education Quarterly Reviews

The Asian Institute of Research *Education Quarterly Reviews* is a peer-reviewed International Journal. The journal covers scholarly articles in the fields of education, linguistics, literature, educational theory, research, and methodologies, curriculum, elementary and secondary education, higher education, foreign language education, teaching and learning, teacher education, education of special groups, and other fields of study related to education. As the journal is Open Access, it ensures high visibility and the increase of citations for all research articles published. The *Education Quarterly Reviews* aims to facilitate scholarly work on recent theoretical and practical aspects of education.

Investigating Lexical Concept and Semantic Representation of Covid-19 in Coronavirus Corpus: A Corpus-Based Study

Elvi Citraresmana[1], Erlina[2], Inu Isnaeni Sidiq[3]

[1] Faculty of Cultural Sciences, Universitas Padjadjaran, Indonesia

Correspondence: Elvi Citraresmana, Faculty of Cultural Sciences, Universitas Padjadjaran, Jl. Soekarno Hatta km 21, Jatinangor-Sumedang, 45363, Indonesia. E-mail: elvi.citraresmana@unpad.ac.id

Abstract

This article discusses the lexical and semantic representation through the collocation that appeared in the Coronavirus Corpus. This research investigates the frequent collocates that appeared together with the node word Corona and find out how those collocates construct the meaning through the linguistic system and conceptual system as they are involved in lexical representation. This research offers a new insight into teaching language using Lexical Concepts and Cognitive Models (Evans, 2009). The data collected are from the Coronavirus corpus by using the corpus-based method. The strength of the keywords and collocate is measured by using Mutual Information (MI). The MI was set in 5; therefore, three lexemes resulted, i.e., cases, patients, and outbreak. The data were analyzed using the lexical concept and cognitive model proposed by Evans (2009). The research results, in general, reveal that the information is coming from the "authorized institution" and "government's representative"; it needs "the official approval or agreement" before publishing to the media, and in the passive form, it describes the foregrounded information and agentless informational assertion. Therefore, the information should be accurate and firm. However, another lexical concept reveals that the information is terrorizing and terrifying; the educated and trained person should also conduct the treatment.

Keywords: Coronavirus, Corpus-Based, Frequency, Lexical Concept, Mutual Information, Semantic Representation

1. Introduction

Since this virus spreads globally, newspapers, magazines, electronic media, and social media have published articles on Coronavirus. The government keeps posting how to deal with or avoid people getting infected by that virus. If we look at the history of this Coronavirus, it was backdated to November 17, 2019. According to South China Morning Post (LiveScience Report, published March 14, 2020) accessed December 18, 2020, they said that this disease started to spread across the globe "when a 55-year-old individual from Hubei Province in China may have been the first person to have contracted COVID-19." (Bryner, 2020). This study investigates the lexical concepts that appeared in the collocates of node words "corona" published in the Coronavirus corpus. The previous research about Coronavirus has not discussed the data based on the Lexical Concept and Cognitive Model (LCCM) analysis. Therefore, this article offers a new perspective on investigating the lexical dan semantic representation

that appeared in the collocates of Corona using Lexical Cognitive Model analysis as a linguistically mediated simulation (Evans, 2009). The present researchers convince that this approach would benefit English learners and English teachers to understand how to interpret the concept of words and understand the meaning.

How is this disease contagious? Or how people are getting infected by this virus? They are the questions that are not treated as the primary focus. Instead, the main focus of this research is the collocates that frequently appear together with the node words Corona in news articles, television, electronic media, and even social media.

The coronavirus corpus was created during this pandemic and was released by May 2020. Creating this Corpus intends to discover what people are saying in online daily papers and magazines in 20 distinctive English-speaking nations (Mark Davies, 2020). His research aims at identifying the "collocates" that appear together with the node words "corona." Furthermore, the collocations which arise will be analyzed based on the corpus method. Finally, this research also describes how those collocations formed semantic representations using Lexical Concept and Cognitive Models (LCCM).

Since this virus spread out around the globe, several linguists have conducted research concerning Covid-19. This pandemic draws attention not only the scientists but also to linguists. The articles discussed Covid-19, among others, are written by Olimat (2020), who investigated whether Jordanian Arabic society used euphemism or dysphemism when discussing the Covid-19 pandemic. He distributed questionnaires to 200 Jordanian respondents concerning the demographic information and open-ended and closed-ended questions. He used the sociolinguistics approach based on Allan and Burridge's theory (1991; 2006) concerning euphemism and dysphemism, together with Warren's model of euphemism (1992) also Lakoff and Johnson's Conceptual Metaphor Theory (1980). His research results show that Jordanians use a different euphemistic techniques in daily Covid-19 conversation, such as metaphor, a shift from Arabic into English, medical terms, and abbreviations, but they hardly use dysphemism.

When Olimat (2020) investigated Covid-19 by distributing a questionnaire to Jordanian respondents, other research based on Sociolinguistic theory conducted by Malaysian researcher Kasdan et al. (2020) observed the terminology Covid-19 Malay Language Terminology Corpus using Socio-terminology theory. Their study aimed at analyzing the coinage of the terms related to Covid-19 in which the community-initiated it. They used the data source from the text released by the Ministry of Health Malaysia from January 16 2020, to May 10, 2020. The terms studied were compared to the DBP database. The results show that DBP already standardizes 67.0% of terms used by the Ministry of Health Malaysia. Even though some words were new, they were still in line with the guideline. However, they were unfavorable among terminologists.

Other research concerning Covid-19 also was conducted by Malaysian researchers. Nor & Zulcafli (2020) investigated how the news of Covid-19 was broadcast in News reports based on Corpus Driven study. They used an online newspaper report about Covid-19 downloaded from the Star online from March 1 to March 31, 2020, as the data source. They collected about 1018 news reports, and 140 themes came up. In addition, they chose 100 top collocates with the MI score minimum set 5 at least by using Antconc software. Their research results show that the collocates reflect fear, anxiety, and uncertainty that most Malaysians feel.

Joharry & Turiman (2020) also investigated the Malaysian public letter to the editor on the Covid-19 pandemic based on a Corpus-assisted discourse study, similar to research conducted by Nor and Zulcafli. Joharri and Turiman used a combination between corpus methods and classic CDA. Their investigation revealed that mostly Malaysian readers showed negative expressions, especially in an emotional way.

Another research article written by Rajandran (2020) discussed his finding of a negative expression in the data. He examined how the Prime Minister of Malaysia and Singapore framed Covid-19 through Metaphors. Rajandran used the Metaphor Identification Procedure and Metaphor Interpretation. The data were taken from the Prime Minister's broadcasts on or to Media in March, April, and May 2020. Using Metaphor Identification Procedure and Metaphor Interpretation shows that both Prime Ministers used a similar metaphor, resulting in COVID-19 IS WAR.

The four studies explained above used the same sources, Malaysia newspapers, whether articles or public letters sent to the editor, show the negative expression concerning of Covid-19 pandemic. However, one of the studies concerning the Covid-19 pandemic has a different perspective. Rapi & Sultan (2020) employed the Positive Discourse Analysis framework. They collected 28 recorded press conference sessions announced by the Indonesian government spokesperson's speech. Their study shows that the government's spokesperson used the strategies such as *"nomination, predication, argumentation, perspectivization and intensification, and mitigation"* to deliver information to the public. The whole discursive strategies are used to improve the public's optimism and build solidarity among the citizens as a moral force to face the pandemic.

On the other hand, primarily Jordanian Arabic society used euphemisms in broadcasting the pandemic situation of COVID-19 (Olimat, 2020). A similar result of what Olimat had done also happened in the study conducted by Rapi & Sultan (2020). According to Rapi & Sultan (2020), the information released by the spokesperson in Indonesia gave a positive result; using discourse strategies, the Indonesian government improved the public's optimism and built solidarity among the citizens.

This research fills the gap between the previous research conducted by Nor & Zulcafli (2020) and Rajandran (2020). Nor and Zucafli used the articles published in the star online news portal. This news portal discussed the issue of Covid-19 in Malaysia. In comparison, this research used the Coronavirus corpus. This Corpus collected the issue of the Coronavirus pandemic in 20 different English-speaking countries.

Nor & Zulcafli (2020) investigation resulted in the collocates appearing with Covid-19. In addition, one of their research results found the noun collocates 'cases', 'outbreak,' and 'patients' as the top three most frequent noun collocates. Our research also found that the noun collocates frequently appeared with the keyword Corona. In contrast with Nor Fariza Mohd. Nor and Zulcafli, our research conducted the study based on Cognitive Model using a corpus-based method during Nor Fariza Mohd. Nor and Zulcafli conducted the research based on Corpus driven study combined with Discourse analysis. This current research is also under the semantic cognitive theory of Rajandran (2020). However, Rajandran used Metaphor Identification Procedure and Metaphor Interpretation, and the researchers analyzed the data using the Lexical Concept and Meaning Construction based on LCCM theory (the abbreviation of LCCM will be applied further).

Fillmore has three main contributions to Linguistic theory. They are case grammar, frame semantics, and construction grammar, as cited in (Hank, 2013). This research focuses on frame semantics initiated by (C. Fillmore, 1975; 1976; 1982a; 1985). We aimed at figuring out the conceptual meaning of lexemes "cases," "outbreak," "patients" that appeared in the Coronavirus corpus. The conceptual meaning is in line with what Fillmore said about 'frame' that the frame is "the system of concepts… to understand the meaning, we have to understand the whole structure in which it fits" (Fillmore, 1982a). To conclude, Fillmore's opinion is that frame semantics focus on analyzing conceptual relations.

Linguists have a variation concept of frame semantics. Among others are, Fillmore & Kay mentioned it as Linguistic System (Fillmore et al., 1988). Evans considers it a Symbolic Unit (Evans, 2009). Goldberg refers to it as Cognitive Construction Grammar (Goldberg, 1995; 2006). At the same time, Langacker tends to use it as Cognitive Grammar for complex symbolic assemblies (Langacker, 1987, 2008). However, all names have the same intention, i.e., finding out the lexical concept and meaning construction. Croft illustrated the anatomy of symbolic units as follows
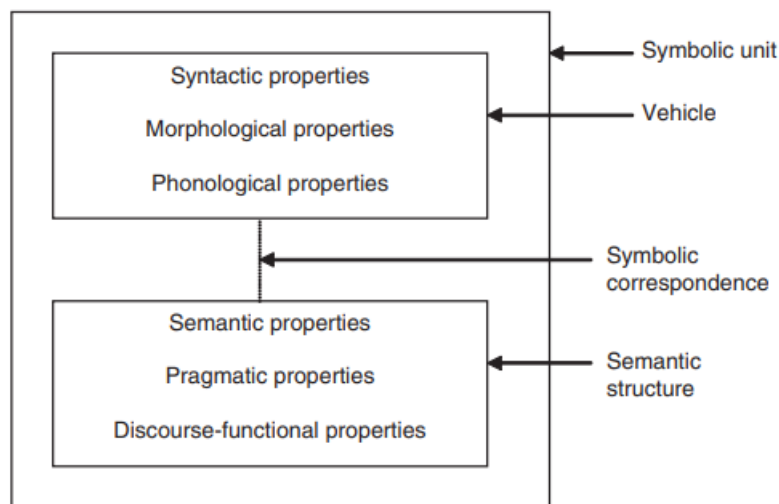
Figure 1: Anatomy Symbolic Units' Croft model (adapted by Evans, 2009, p. 95).

Evans (2009) discusses the structure of symbolic units based on Croft's model above as follows

    a. Vehicle: "France"
          Lexical concept: [FRANCE]
    b. Vehicle: "NP *kick*FINITE *the bucket*"
          Lexical concept: [AN ANIMATE ENTITIES DIES]
    c. Vehicle: "NP FINITE VERB NP NP"
          Lexical concept: [THING X CAUSES THING Y TO RECEIVE THING Z]

Evans (2009) proposed the Lexical Concept Cognitive Model (LCCM theory) as a symbolic unit. This LCCM is the representation type held to populate the linguistic system. According to Evans, a symbolic unit is a bipolar assembly of phonological content represented by a vehicle. Thus, the lexical concept forms the semantic structure, while the vehicle is phonetically overt.

The lexical concept is formed by paired closed-class vehicles and paired open-class vehicles. The paired closed-class vehicle results encode the linguistic content and paired open-class vehicles could form as encoding linguistic content and provide access site to conceptual content. Therefore, the cognitive model proposed by Evans (2009) is considered the appropriate theory to analyze the data.

Collocation is concerned with rehashed co-occurrence of words and writing (Brezina, 2019). To investigate the distribution of collocation, we observed frequency (F) and Mutual Information (MI). This collocation could appear in specific texts. They could occur at a very high frequency, appear in other texts at a low frequency, or are absent. The bias of frequency of words happens probably because of the text genre, or they appear in a limited genre.

Sinclair and Stubbs proposed collocation as simplified by (Lindquist, 2009, p. 57) as "the relation between a word or individual word-forms which frequently co-occur with it." The term collocation was motivated by the educationalist (Palmer, 1933). Palmer classified collocations based on several patterns. The term collocation is also defined by British linguist Firth as cited in Lindquist (2009) that the word's meaning is frequently influenced by the phrase accompanied by it.

There are two types of collocations, window collocations and adjacent collocations. Both have different methods. For example, in window collocations, they search the frequent words that appeared to the left and the right keywords in 4 to 5. This collocation aligns with Firth and Sinclair's proposal (cited in Lindquist, 2009); see Stubbs, 2002. However, Sinclair considered using the strength of the measures in a varied way based on the statistical measure considered appropriate.
The second type of collocation was in line with Palmer, and it is known as the Palmerian point of view (cited in Lindquist, 2009). This type of collocation search for the frequent collocates appeared immediately to the left or

right of the keywords. According to Lindquist, this adjacent collocation method is closer to the investigation based on a linguistic approach. It is different from window collocation, which is quantitative rather than qualitative. However, this research uses the combination of window and adjacent collocation.

This research investigates the frequency of collocates that appeared with the node words Corona. The MI score adopts window collocation to measure the strength between node words and the collocates. In searching for the frequency of words, this research also adopted the perspective of adjacent collocation by utilizing the measurement five for the MI score. Another tool to be used is frequency. Frequency also is an essential part of corpus linguistics. Furthermore, frequency helps the researchers handle massive data. The researchers will be able to identify the data by observing the frequencies of the occurrence of the words as the frequent words or frequently used word combinations (see Lindquist, 2009).

The language research can observe the concordance that appeared in the corpus tools. The researchers need to read and analyze concordances to find semantic or grammatical information. The researcher will quickly get the information by using concordance tools rather than reading the whole text, and then they decide on the data source.

This research uses Antconc software as third-generation concordance (McEnery & Hardie, 2011). Concordance is another helpful tool to see the span of a window of the text (Palmer; Firth as cited in Lindquist, 2009). First, collocates of "corona" were collected based on the frequency in the Coronavirus corpus. Next, those collocates are chosen based on MI. Finally, we set the score of MI to a maximum of 5. Those data sets are considered corpora. Next, we transferred those corpora into Antconc software. Through the Antconc software, we have the collocates of lexemes "cases," "outbreak," and "patients. Those collocates were collected based on the cluster/N-grams and analyzed in the LCCM theory scheme.

2. Method

This research is a corpus-based study in line with Corpus Linguistics. The data were collected using descriptive statistics through F and MI. The data were analyzed in a descriptive method. Since this pandemic, Brigham Young University created Coronavirus Corpus, released in May 2020. They collected the data from online newspapers and magazines from 20 different English-Speaking countries. The keyword chosen in this research is Corona since this research aims to find out what those people in 20 different English-speaking countries are saying about this virus. The collocates are selected by using F and MI, as mentioned earlier.

Coronavirus corpus size is currently 747 million words. Corpus allows the researchers to search for the frequent words that appeared in online newspapers and magazines in 20 English-speaking countries during the pandemic crisis. First, the researchers collected the data from the Coronavirus corpus based on the frequency. Next, the closeness of collocates and the node words is observed based on MI. Finally, the MI score is set at a maximum of 5.

Those purposive data then were transferred into Antconc software to find the span of the collocation. Furthermore, collocates are helpful to collect the word type of words that appeared in those collocates. Through F and MI, they result in three lexemes "cases," "patients," and "outbreak." Finally, we analyze the data based on lexical and semantic representation, which aligns with the Cognitive Linguistic study.

This Corpus has many benefits for the researchers since Corpus is the only comfortable way to search for worldwide phenomena without conducting field research. Based on our observations and experiences, the announcement constantly caused people's fear in the first semester. The frequent words the spokesman, the newspapers, or any media announce are the increasing numbers of infected people and people who recovered from the disease. Other frequent words during this pandemic are virus, pandemic cases, crisis, patients, positive, warriors, outbreak, disease, and infection. The frequencies are 9544, 1695, 1366, 1083, 988, 810, 731, 675, 549, 398 tokens respectively.

3. Results

In the first place, the researchers observed the collocates appeared with the word(s) Corona followed by All (not a specific category).

Table 1: The Frequency of Collocates words Corona

NO.	NODE WORDS "CORONA"	COLLOCATES	F
1		Virus	9595
2		Pandemic	1713
3		Cases	1380
4		Crisis	1097
5		Patients	999
6		Positive	827
7		Warriors	738
8		Outbreak	677
9		Disease	552
10		Infection	400

Source: Coronavirus corpus accessed on Dec. 13, 2020

In the second place, the researchers observed the collocates sorted by noun category. The result is shown in Table 2 below.

Table 2: The top 10 noun Collocates of CORONA

NO	Noun Collocates	F	MI
1	Virus	9600	8.05
2	Cases	1381	5.00
3	Crisis	1101	6.10
4	Patients	1003	5.85
5	Warriors	738	10.54
6	Outbreak	677	5.07
7	Disease	552	5.15
8	Spread	526	3.42
9	Infection	425	4.84
10	Number	378	3.07

Source: Coronavirus corpus accessed on Dec. 13, 2020

Table 2 above shows us that noun collocates appeared on the right side of node word corona; they form into NP such as 'Coronavirus,' 'Corona Cases,' and so on. However, to find out the strength of both collocations needs the statistical measures of MI. As we can see, according to Latham & Roudi (2009), "high MI indicates a large reduction in uncertainty and low MI indicates small reduction and zero MI between the two random variables means the variables are independent." Therefore, it takes the consequence that this collocate does not share mutual information due to a considerable reduction in uncertainty.

As mentioned earlier, this research has set the MI score to a maximum of 5 causes the results that the collocates analyzed are "cases," "patients," and "outbreak." Therefore, the discussion is limited to 3 data considered representative data.

To find the collocates of lexemes 'cases', 'outbreak,' and 'patients' collected from the Coronavirus corpus, we transferred the data into Antconc software. From table 3 below, we can see that the collocates of cases are the word 'xa' as amount as F=342. However, this word cannot be understood and is considered fragmented, and we have to exclude the word 'xa' as our data. Then, the paired closed-class vehicles appeared as collocating cases such as the preposition 'of' and determiner 'the.' The paired with open-class vehicles also appeared, such as 'corona,' 'confirmed,' 'suspected.'

Table 3: The collocates of cases

Rank	Freq	Freq (L)	Freq (R)	Stat	Collocates
1	342	95	247	4.46212	Xa
2	96	12	84	4.09276	Of
3	85	2	83	4.45775	Corona
4	16	13	3	2.15024	The
5	16	14	2	4.18587	Confirmed
6	14	14	0	3.99322	Has
7	12	11	1	4.63333	Suspected
8	11	10	1	4.38226	Two
9	10	10	0	2.82972	To
10	10	8	2	1.94708	In

Source: Coronavirus corpus accessed on Dec. 13, 2020

From Antconc software (see table 4 below), we can see the collocates of the outbreak is the paired closed-class vehicles. They are determiners 'the' followed by the preposition 'of,' 'to,' 'in,' 'since,' 'after,' 'with,' 'over,' and 'due.' We limited the frequency not below 10. There are no paired with open-class vehicles. As in line with Evans's lexical concept, this paired close-class will encode linguistic content. As in the data example, "the outbreak of /…/.

Table 4: The collocates of the outbreak

Rank	Freq	Freq (L)	Freq (R)	Stat	Collocates
1	338	88	250	4.65126	Xa
2	79	79	0	3.68581	The
3	102	18	84	4.30468	Of
4	15	15	0	3.22598	To
5	12	11	1	2.07178	In
6	7	7	0	4.67665	Since
7	7	7	0	4.48400	Due
8	6	6	0	3.80218	After
9	5	5	0	4.41361	Over
10	4	4	0	3.86929	With

Source: Coronavirus corpus accessed on Dec. 13, 2020

The collocates patients (see table 5 below) are followed by paired open-class vehicle 'corona' and the paired closed-class preposition 'of,' 'in,' 'for,' 'to,' determiner 'the,' and conjunction. That paired closed-class which encodes linguistic content, will be analyzed in detail in the next section.

Table 5: The collocates of patients

Rank	Freq	Freq (L)	Freq (R)	Stat	Collocates
1	305	161	144	4.44864	Xa
2	98	60	38	4.38100	Corona
3	45	25	20	3.73775	Of
4	37	22	15	3.00519	The
5	27	5	22	3.21980	In
6	17	11	6	3.09079	And
7	14	10	4	3.11554	For
8	13	8	5	2.70377	To
9	13	4	9	3.45223	Are
10	11	6	5	2.81069	That

Source: Coronavirus corpus accessed on Dec. 13, 2020

We analyze the collocation based on the above collocates findings that appeared in table 3, table 4, and 5 to find the lexical concept and semantic representation using the LCCM theory. The schematic semantic content itself refers to the Cambridge Dictionary.

4. Discussion

4.1. The Lexical Concept of Cases and Its Semantic Representation

The verb collocates "confirmed."

A. The perfect tenses
    (1) India too **has confirmed** three cases of Coronavirus in Kerala
    (2) Research (ICMR) **has confirmed** 4 positive cases of Corona in Karnataka, and within minutes of ICMR stating the positive case in the state
    (3) NAN reports that Nigeria currently **has two confirmed** cases of Corona Virus; the first was reported on February 27
    (4) The government **has confirmed** eight more cases of Coronavirus

Table 6.1: Schematic content associated with closed-class vehicles "has confirmed"

Closed-class vehicles	Schematic semantic content
Lexical class: verb (has/have confirmed)	Designates an entity as an event that has been started in the past and is still in progress.
Lexical class: noun (India; Research (ICMR); Nigeria; The Government; cases; Coronavirus)	Designates an entity as an object (as one possibility)
Grammatical relation: subject (visiting countries; India; Research (ICMR); Nigeria: The government)	Designates entity being the primary or focal entity in a designated relationship
Grammatical relation: object (cases; Coronavirus)	The secondary entity in a designated relationship
Active voice: through the verb form	Point of view being situated at the agent
Declarative word order	Speakers know the situation is to be true and assert it to the hearer

Table 6.2: Rich content associated with open-class vehicles "has confirmed"

Open-class vehicles	Rich semantic content
India	Proper place
Research (ICMR)	An institution that conducts the research
Nigeria	Proper place
The government	Proper name functions as a collective noun referring to people who conduct the rules of the country

The schematic content associated with the closed class and rich content associated with open class vehicles "has confirmed" shows [ANNOUNCEMENT OF EXTRAORDINARY EVENTS].

This announcement was launched by the authorized institution, organization, and government representatives. Therefore, it shows [LEGAL ANNOUNCEMENT] lexical concept. Furthermore, this second lexical concept brings forth the third lexical concept [INFORMATIONAL ASSERTION]. Those lexical concepts encode linguistic content that "has confirmed" pragmatically refers to the announcement concerning Covid-19. The participants involved are those who have the authorization. Therefore, the announcement is legal as the information assertion. To conclude, the semantic representation revealed from this lexical concept is [N/NP1(the authorized institution) has confirmed (legal information) N/NP2(information assertion)].

    B. Simple past tense
    (5) with cases of Coronavirus or Cove in 19 now **confirmed** in New York, Vermont, and New
    (6) Institute also **confirmed** that there are no reported cases of Corona Virus in Sri Lanka so far.

(7) Now 55 **confirmed** cases of Coronavirus in San Diego County.
(8) Shares # Telangana CM KCR **confirmed** 10 new cases of Corona positive in a single day at the press meet he called for.

Table 6.3: Schematic content associated with closed-class vehicles "confirmed"

Closed-class vehicles	Schematic semantic content
Lexical class: verb (confirmed)	Designates an entity as an event that happened in the past; it proved accurate, and an entity was approved officially by formal agreement.
Lexical class: prep (in)	Designates entity is inside a container, place, or area.
Grammatical relation: subject (Cases of Coronavirus or Cove in 19); Institute; 55; Telanggana CM KCR)	Designates entity being the primary or focal entity in a designated relationship
Grammatical relation: object (cases of Coronavirus; that there are no reported cases of Corona Virus; cases of Coronavirus; 10 new cases of Corona positive)	The secondary entity in a designated relationship
Active voice: through the verb form	Point of view being situated at the agent
Declarative word order	Speakers know the situation is to be true and assert it to the hearer

Table 6.4: Rich content associated with open-class vehicles "confirmed"

Open-class vehicles	Rich semantic content
Institute	Proper name functions as a collective noun referring to academics, such as research and educational matters.
55	Designates the amount or number of the entities
Telangana CM KCR	A proper name refers to a person who has authority in India.

The schematic content associated with the closed class and rich semantic content of vehicles "confirmed" shows the lexical concept of [FORMAL APPROVAL]. Since the entity is a formal approval, the authorized institution and the government's representative are involved. So, the lexical concept revealed here is [AUTHORISATION AGREEMENT]. Those agreements bring forth [INFORMATIONAL ASSERTION] lexical concept. To conclude, the semantic representation revealed from this lexical concept is [N/NP1(authorized institution) confirmed (formal approval) (authorization agreement) N/NP2(informational assertion)].

C. Passive Sentence
(9) About 525 cases **are confirmed**
(10) postpone the holiday, despite there **being confirmed** cases of Coronavirus in Delhi," she told me.

Table 6.5: Schematic content associated with closed-class vehicles "are confirmed"

Closed-class vehicles	Schematic semantic content
Lexical class: verb (are confirmed)	Designates an entity as an event that happened in the past; this entity proved true and was approved officially by formal agreement.
Lexical class: prep (in)	Designates something or someone is inside a container, place, or area.
Grammatical relation: subject (525 cases; despite there)	Designates entity being the primary or focal entity in a designated relationship
Grammatical relation: complement (cases of Coronavirus)	The secondary entity in a designated relationship

Passive voice: through the verb form	Designates entities occurred not from the point of view of the agent
Declarative word order	Speakers know the situation is to be true and assert it to the hearer

The lexical concept of the vehicle "are confirmed" shows a similar semantic representation with vehicles "confirmed." However, the information occurred not from the point of the agent. Therefore, it reveals the [AGENTLESS INFORMATIONAL ASSERTION] lexical concept. To conclude, the semantic representation of vehicles "are confirmed" in passive voice is [N/NP(topics) are confirmed(formal approval)] the writers did not insert the agents involved.

The verb collocates "have been reported"

    (11) Cases of Coronavirus **have also been reported** from countries including Thailand, Vietnam, Singapore

    (12) to travelers from China or any country where cases of Coronavirus **have been reported** to observe self-quarantine on arrival in Lagos.

    (13) Dr. Zafar Mirza said that 4,446 suspected cases of Coronavirus **have been reported** in Pakistan

    (14) Two suspected cases of Corona Virus in Akwa Ibom State **have been reported** negative.

    (15) 2020, there **have been only two reported** cases of Coronavirus in Michigan.

Table 6.6: Schematic content associated with closed-class vehicles "have been reported"

Closed-class vehicles	Schematic semantic content
Lexical class: verb (have been reported)	A description of an entity's designation. Designates an entity to prove to be true. Designates an entity being approved officially by formal agreement.
Lexical class: prep (in)	Designates something or someone is inside a container, place, or area.
Lexical class: prep (from)	Designates the place where the entities start.
Lexical class: prep (to)	Designates the action described in the infinitive that will happen later.
Grammatical relation: subject (Cases of Coronavirus; any country where cases of Coronavirus; 4,446 suspected cases of Coronavirus; Two suspected cases of Corona Virus)	Designates entity being the primary or focal entity in a designated relationship
Grammatical relation: complement (cases of Coronavirus: to observe self-quarantine on arrival)	The secondary entity in a designated relationship
Passive voice: through the verb form	Designates entities occurred not from the point of view of the agent
Declarative word order	Speakers know the situation is to be true and assert it to the hearer

The schematic content associated with the closed class vehicles "have been reported" shows [OFFICIAL APPROVAL] and [OFFICIAL AGREEMENT] lexical concepts. Since the vehicle is in a passive form, there is not any agent described. However, the event is foreground to highlight the information. As a result, the vehicles "have been reported" show [FOREGROUND INFORMATION] lexical concept. To conclude, the semantic representation of vehicle "have been reported" are [N/NP1(foreground information) have been reported (official approval/agreement) to-inf(action) in(proper place)] and [N/NP1(foreground information) have been reported(official approval/agreement) adj (state of entity's condition)]; [There(expletive) have been reported(official approval/agreement) NP1(psychological subject)].

The verb collocates recorded

    (16) announced on Saturday that it **has recorded** 10 cases of Coronavirus.

(17) The United States **has recorded** 122,000 cases of Coronavirus so far, the largest number of infections in one country.

(18) announced on Monday that it **had recorded** 22 cases of Coronavirus during the past 24 hours in the region.

(19) Kurdistan Region **has recorded** 161 cases of Coronavirus so far,

Table 6.7: Schematic content associated with closed-class vehicles "has recorded"

Closed-class vehicles	Schematic semantic content
Lexical class: verb (has recorded)	Designates devices used to record; the entity does an activity using electronic tools.
Lexical class: prep (during)	Designates entity occurs from the beginning to the end of a particular period.
Grammatical relation: subject (Cases of Coronavirus; any country where cases of Coronavirus; 4,446 suspected cases of Coronavirus; Two suspected cases of Corona Virus)	Designates entity being the primary or focal entity in a designated relationship
Grammatical relation: complement (cases of Coronavirus: to observe self-quarantine on arrival)	The secondary entity in a designated relationship
Declarative word order	Speakers know the situation is to be true and assert it to the hearer

The schematic content associated with the closed class vehicles "has recorded" shows the entity has done something using the electronic tools. The agent does the activity at a time, containing formal and accurate information. The speaker knows the situation is actual asserts it to the hearer. As a result, that schematic content reveals [ACCURATE INFORMATION BY TAKING NOTES] lexical concepts. To conclude, those lexical concepts show a semantic representation as follows [N/NP1(government's representative) has recorded (take notes activity) NP2(accurate information)].

The adjective collocates confirmed

(20) 2020, Italy reported only three **confirmed** cases of the Coronavirus.

(21) These **confirmed** cases of Coronavirus not cooperating with health care officials/police should be handled under the National Security Act

(22) blamed for the sudden spike of **confirmed** positive cases of Corona Virus Disease -19 cases in Malaysia.

(23) **confirmed** cases of Corona Virus in New York are of Asian descent

Table 6.8: Schematic content associated with closed-class vehicles "confirmed"

Closed-class vehicles	Schematic semantic content
Lexical class: verb (reported)	Designates the entity's activity in informing something to someone.
Lexical class: adj (confirmed)	Designates entity firmly fixed.
Lexical class: prep (in)	Designates something or someone is inside a container, place, or area.
Lexical class: prep (of)	Designates entity's possession or belonging.
Lexical class: prep (under)	Designates entity is experiencing something.
Grammatical relation: subject (Italy, cases of Coronavirus)	Designates entity being the primary or focal entity in a designated relationship
Grammatical relation: object (only three confirmed cases of Coronavirus; the National Security Act; Asian descent)	The secondary entity in a designated relationship
Declarative word order	Speakers know the situation is to be true and assert it to the hearer

Table 6.9: Rich content associated with open-class vehicles

Open-class vehicles	Rich semantic content
Italy	Proper name functions as a collective noun referring to the place and people who have the authority.
Asian descent	Proper nouns to designate the specific races.

The closed-class vehicles "confirmed" schematic content shows the lexical concept [FIRM INFORMATION]. Based on the rich content associated with open-class, the participant who announces the event refers to the government's representative or the authorized institution. Therefore, it reveals the [AUTHORIZED INSTITUTION INVOLVED] lexical concept. The second participants belong to the patients. As a result, it shows [PATIENTS' INFORMATION] lexical concept. To conclude, that lexical concept reveals a semantic representation [confirmed(adj)(firm information) (patients') (authorized institution involved)].

The adjective collocates suspected
      (24) Coronavirus in Sri Lanka Two **suspected** cases of Coronavirus in Sri Lanka # Written by Staff Writer #January 25n 2020
      (25) For **suspected** serious cases of the Corona, employers are strongly advised to notify the public health department
      (26) for doctors and medical staff to treat **suspected** cases of Corona: An allowance!
      (27) by the Institute of Infectious Diseases with **suspected** cases of Corona Virus infections being reported from Sri Lanka.

Table 6.10: Schematic content associated with closed-class vehicles "suspected"

Closed-class vehicles	Schematic semantic content
Lexical class: adj (suspected)	Designates the information that is believed to be true.
Lexical class: noun (two)	Designates number or amount
Lexical class: noun (Sri Lanka)	Designates proper place
Lexical class: noun phrase (cases of Coronavirus; doctors and medical staff; the Institute of Infectious Diseases; the public health department)	Designates to describe the entity as modified by the adjective
Lexical class: to inf (to notify)	Designates to inform the entity about something
Lexical class: prep (in)	Designates something or someone is inside a container, place, or area.
Lexical class: prep (of)	Designates entity's possession or belonging.
Lexical class: prep (from)	Designates the place where the entities start.
Declarative word order	Speakers know the situation is to be true and assert it to the hearer

The schematic content associated with the closed-class vehicles "suspected" shows the entity's information that is believed to be true. As a result, it reveals [POTENTIAL STATUS] lexical concept. That lexical concept shows the semantic representation [N(number) suspected(adj)(potential status) NP(the caused)]; [to-inf(action) suspected(potential status) NP(the caused)].
The adjective collocates positive
      (28) several cases of corona-**positive** patients have been found
      (29) Meanwhile, the total cases of Corona **positive** have crossed three thousand in the country.
      (30) We have received confirmation of two more cases of Corona; the two have **tested positive** as a result of coming into contact with
      (31) from #Mohali was **tested positive** for Corona.

Table 6.11: Schematic content associated with closed-class vehicles "positive"

Closed-class vehicles	Schematic semantic content

Lexical class: adj (positive)	Designates the entity's test result and is approved officially by formal test.
Lexical class: noun (patients)	Designates entity's condition
Lexical class: noun (Corona)	Designates proper name
Lexical class: noun phrase (several cases of Corona; the total cases of Corona; the two; as a result of; three thousand; the country)	Designates to describe the entities as modified by the adjective
Lexical class: a passive form of the verb phrase (have been found)	Designates to describe the agent finds the entity
Lexical class: a passive form of the verb phrase (was tested; have been found)	Designates the event, not from the perspective of the agent
Lexical class: verb phrase (have tested)	Designates a discovery of the entity's condition
Lexical class: verb phrase (have crossed)	Designates the entity goes one side to another side
Lexical class: prep (in)	Designates something or someone is inside a container, place, or area.
Lexical class: prep (for)	Designates to show the entity's purpose or the entity's intention
Declarative word order	Speakers know the situation is to be true and assert it to the hearer

The schematic content associated with the closed-class vehicles "positive" refers to the test's result, examined by an official institution. Therefore, it reveals [MEDICAL TEST RESULT], [OFFICIAL INFORMATION], and [NEGATIVE EVALUATION] lexical concepts. That lexical concepts show semantic representation [NP V(event)(active form) Adj(medical result)], [NP V(event)(passive form)(official information], [N V(event)(passive form) (negative evaluation) for(prep) N(result)].

4.2. The Lexical Concept of Outbreak and Its Semantic Representation

The noun collocates outbreak

(32) The minister said more than 70 billion on a daily basis since before **the outbreak of Corona Virus**

(33) **After the outbreak of Corona**, Contagion started trending and became the' Most Demand Film' in the

(34) **Another outbreak of Corona Virus** was suspected at Zoom HQ in China as many workers there were found

(35) in a bid to help combat **the deadly outbreak of the Coronavirus** in mainland China.

(36) The Islands are **in fear of an outbreak of Corona,** so Pacifica is no brained [sin] to cancel

Table 6.12: Schematic content associated with closed-class vehicles "outbreak"

Closed-class vehicles	Schematic semantic content
Lexical class: noun (outbreak)	Designates the dangerous, unpleasant entity (usually disease) begins all of a sudden
Lexical class: determiner, pronoun (another)	Designates additional or one more entity
Lexical class: adj (deadly)	Designates to describe the entity which caused to death
Lexical class: prepositional phrase (in fear of)	Designates the entity's feelings
Lexical class: prep (before)	Designates to describe the agent finds the entity
Lexical class: a passive form of the verb phrase (was tested)	Designates the action conducted by the agent
Lexical class: verb phrase (have tested)	Designates a discovery of the entity's condition
Lexical class: verb phrase (have crossed)	Designates the entity goes one side to another side
Lexical class: prep (in)	Designates something or someone is inside a container, place, or area.

| Lexical class: prep (for) | Designates to show the entity's purpose or the entity's intention |
| Declarative word order | Speakers know the situation is to be true and assert it to the hearer |

The schematic content associated with the closed-class vehicles "outbreak" refers to disease, dangerous events that occurred without notice in advance, violence, caused to death. Therefore, it reveals [QUANTIFIER] lexical concept such in "Another outbreak of Corona Virus /…/" see (data 35), [TERRORIZING] lexical concept such in "the deadly outbreak of Coronavirus" see (data 36), [TERRIFYING EVENT] lexical concepts such in "in fear of an outbreak of Corona" see (data 37).

4.3. The Lexical Concept of Patients and Its Semantic Representation

The noun collocates treatment
>   (37) first prototype of hospital isolation coach for the **treatment** of corona patients
>   (38) also ensure complete guidance to Pakistani doctors regarding the **treatment** of corona patients in Punjab.
>   (39) to the need for the staff providing **treatment** to the corona patients.
>   (40) No hospital can refuse to admit patients for corona **treatment,** and if anyone is found doing it, legal action will be taken

Table 6.13: Schematic content associated with closed-class vehicle "treatment"

Closed-class vehicles	Schematic semantic content
Lexical class: noun (treatment)	Designates the way the entity is considered and examined.
Lexical class: determiner (the)	Designates to particular things, people, places
Lexical class: preposition (for)	Designates the intention of someone to give something.
Lexical class: preposition (to)	Designates the action will happen later.
Lexical class: preposition (of)	Designates to show possession, belonging, or origin.
Lexical class: preposition (in)	Designates something or someone is inside a container, place, or area.
Lexical class: noun phrase (the treatment of corona patients)	Designates how the entity is examined concerning medical care.
Declarative word order	Speakers know the situation is to be true and assert it to the hearer

The schematic contents associated with closed-class vehicles "treatment" refer to medical treatment due to Coronavirus; this medical treatment was done by someone who had a medical, educational background. Therefore, it reveals [EDUCATIONAL NEEDS] lexical concepts such as "/…/ coach for the treatment of Corona patient" see (data 38), and "also ensure complete guidance to Pakistani doctors regarding the treatment of corona patients in Punjab" see (data 39). Furthermore, the lexical concept [EXAMINATION SUPPLY] such in "to the need for the staff providing treatment to the corona patients" see (data 40) and "No hospital can refuse to admit patients for corona treatment**,** and if anyone is found doing it, legal action will be taken" see data (41).

This coronavirus corpus is data collected from 20 different English countries. Therefore, we can conclude that this Coronavirus corpus represents the 20 English countries. Furthermore, through F and MI, the three words "cases," "patients," and "outbreaks" most frequently occur in this Corpus. Therefore, this LCCM theory has demonstrated significant feasibility in analyzing collocation cases, patients, and outbreaks. Through vehicles "has confirmed," "confirmed," "are confirmed," "have been reported," "recorded," "confirmed," "suspected," "positive," "outbreak," and "treatment," it reveals that those vehicles result in a semantic representation that all the information is coming from "the authorized institution," "the official approval or official agreement," and "the government's representatives". In the passive form, the information of the entity's condition is foregrounded, and they are agentless informational assertions. The information content is accurate and firm and describes the potential status and medical evaluation test result. In other vehicles, "outbreak" and "treatment" reveal a semantic

representation of "quantifier," "terrorizing," and "terrifying events"; it also describes the "educational needs" and "examination supply."

Through LCCM theory, the current researchers have a very fruitful insight into revealing the meaning construction of the collocation of "cases," "patients," and "outbreak" lexical concepts. Since the previous research discussed Coronavirus and Covid-19 from Semantic Cognitive through Metaphor; Metaphor, Nominalization, Appraisal from the perspective of Critical Media Discourse Analysis and Corpus Driven from the perspective of Critical Discourse Analysis.

Acknowledgements

References

Brezina, V. (2019). Collocation Graphs and Networks: Selected Applications. In *Lexical Collocation Analysis: Advances and Application* (p. 59). Springer Nature.

Bryner, J. (2020). *1st known Case of Coronavirus traced Back to November in China*. Live Science. Retrieved December 18, 2021. https://www.livescience.com/first-case-coronavirus-found.html

Evans, V. (2009). *How Words Mean: Lexical Concepts, Cognitive Models, Meaning Construction*. Oxford University Press.

Fillmore, C. (1975). An Alternative to Checklist Theory of Meanings. *Papers from the First Annual Meeting of the Berkeley Linguistics Society. California.*

Fillmore, C. J. (1976). *Frame Semantics and the Nature of language." In Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech.*

Fillmore, C. J. (1982). *Frame Semantics: Linguistics in the Morning*. Hanshin.

Fillmore, C. J. (1985). Frames and the Semantics of Understanding. *Quaderni Di Semantica*, *6*(2).

Goldberg, A. . (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Cambridge University Press.

Goldberg, A. . (2006). *Construction at Work*. Oxford University Press.

Hank, P. (2013). *Lexical Analysis: Norms and Exploitations*. The MIT Press.

Joharry, S. A., & Turiman, S. (2020). Examining Malaysian Public Letters to Editor on COVID-19 Pandemic: A Corpus-Assisted Discourse Analysis. *GEMA Online Journal of Language Studies*, *20*(3), 242–260. https://doi.org/10.17576/gema-2020-2003-14

Kasdan, J., Baharuddin, R., & Shamsuri, A. S. (2020). Covid-19 dalam Korpus Peristilahan Bahasa Melayu: Analisis Sosioterminologi (Covid-19 in the Corpus Of Malay Terminology: A Socio-terminological Analysis). *GEMA Online® Journal of Language Studies*, *20*(3), 221–241. https://doi.org/10.17576/gema-2020-2003-13

Langacker, R. W. (2008). Cognitive Grammar: A Basic Introduction. In *Cognitive Grammar: A Basic Introduction*. https://doi.org/10.1093/acprof:oso/9780195331967.001.0001

Latham, P. E., & Roudi, Y. (2009). Mutual information. *Scholarpedia*, *4*(1), 1658. https://doi.org/10.4249/SCHOLARPEDIA.1658

Lindquist, H. (2009). *Corpus Linguistics and the Description of Linguistics*. Edinburgh University Press Ltd.

Mark Davies. (2020). *The Corona Virus*. Corpus Data. Retrieved December 23, 2021. https://www.corpusdata.org/

McEnery, T., & Hardie, A. (2011). *Corpus Linguistics : Method, Theory and Practice*. Cambridge University Press.

Nor, N. F. M., & Zulcafli, A. S. (2020). Corpus Driven Analysis of News Reports about Covid-19 in a Malaysian Online Newspaper. *GEMA Online Journal of Language Studies*, *20*(3), 199–220. https://doi.org/10.17576/gema-2020-2003-12

Olimat, S. N. (2020). COVID-19 pandemic: Euphemism and dysphemism in Jordanian Arabic. *GEMA Online Journal of Language Studies*, *20*(3), 268–290. https://doi.org/10.17576/gema-2020-2003-16

Palmer, H. . (1933). *Second Interim Report on English Collocations*. Kaitakusha.

Rajandran, K. (2020). 'A Long Battle Ahead': Malaysian and Singaporean Prime Ministers Employ War Metaphors for COVID-19. *GEMA Online Journal of Language Studies*, *20*(3), 261–267. https://doi.org/10.17576/gema-2020-2003-15

Rapi, M., & Sultan. (2020). Positive Discourse Analysis of the Indonesian Government Spokesperson's Discursive Strategies during the Covid-19 Pandemic. *GEMA Online ® Journal of Language Studies*, *20*(4). https://doi.org/10.17576/gema-2020-2004-14