Comparative Collocation(al) Analysis between Two Malaysian Online Newspapers: A Case Study of the 2020 Sabah Election Day

SITI AEISHA JOHARRY
Akademi Pengajian Bahasa
Universiti Teknologi MARA (UiTM) Shah Alam, Malaysia
aeisha@uitm.edu.my

SYAMIMI TURIMAN
Akademi Pengajian Bahasa
Universiti Teknologi MARA (UiTM) Shah Alam, Malaysia

MAIZURA MOHD NOOR
Akademi Pengajian Bahasa
Universiti Teknologi MARA (UiTM) Shah Alam, Malaysia

ABSTRACT

The cases of COVID-19 in Malaysia recorded a significant spike after the Sabah State Election, which was held on 26 September 2020. This paper aims to examine news reporting strategies before, during, and after the Sabah Election Day in Malaysia. Related articles from two online local newspapers (one mainstream and one alternative) are compiled into a corpus for this purpose and compared against each other. Using the corpus-assisted discourse approach, #LancsBox6.0 was used to compare statistically significant keywords and their collocates to analyse differences in the way the election was portrayed in both media channels. Findings revealed that both newspapers described news in relation to the number of COVID-19 cases in proximity to reporting of the state election. Arguably, reporting of events is (slightly) different and therefore has repercussions on the way news is represented in mainstream and alternative media. For example, although both online sources reported the significance of the Sabah state election to the rise in COVID-19 cases in Malaysia, The Star used more explicit referencing (e.g., is from Sabah; cases in Sabah) and had a more speculative type of reporting as some of the collocations were found to be from politicians' verbatim that indicated their weak arguments while Malaysiakini provided reports that proved the elections to have had a spill-over effect on the total number of COVID-19 cases in Malaysia. Collocation networks also proved to show differences in the way words are used in language and discourse and can be further explored in future studies.

Keywords: COVID-19 collocation; Malaysian newspapers; #LancsBox6.0; News reporting; Sabah elections

INTRODUCTION

Sabah – an eastern Malaysian state, was forced to hold its election on 26 September 2020, in the midst of the COVID-19 pandemic, after the Warisan-led state government lost its majority due to the defection of its assemblymen. During the COVID-19 pandemic, the Internet brought about many changes in the dissemination of news and information, which in turn provides a wealth of research opportunities. As Lai (2019) contends, "language can be dynamically shaped and reshaped to construct or to manipulate social realities" (p. 545), and among the linguistic devices available for the analysis of news reporting would be collocation analysis as it is considered useful; the presence of recurring patterns can identify specific indicators around words immediately adjacent to each other (Baker et al., 2013).

Collocation, as Firth (1957, p. 2) points out, is "the habitual co-occurrence of words" that is measured in corpus linguistics using a number of statistics (Brezina, 2018). In the English language, words co-occur together to form meanings and are representative of the interlocutor's language and thought processes, and in turn, "[i]nvestigating collocations thus creates an opportunity for looking into the fundamental fabric of text or speech through the lens of connection and association between words" (Brezina, 2018, p. 59). Typical statistical measures that are used in a collocation analysis include t-score and MI-value, where the latter discriminates between infrequent but strongly associated words (compared to the former, which emphasises the frequency of word associations rather than the strength of associations).

Although collocation analysis is nothing new and is relatively able to highlight word associations that could tell us more about a collection of texts - for example, see Prihantoro (2022) for the rising collocates 'renewable energy' over time, "*shared* collocates" on the other hand are still less reported in the literature (exceptions include, Jaafar, 2022; Joharry & Turiman, 2020;). Shared collocates can be described as "[s]imple collocation graphs and collocation networks [that] show association and cross-association between words in language and discourse and can thus be used in a range of areas of linguistic and social research" (Brezina, 2018, p. 59). For the purpose of this study, a collocation analysis of shared collocates is therefore useful in order to discern between the different ways in which newspapers report about a common event, such as the 2020 Sabah state election in Malaysia. More specifically, the research objectives of the present study are to:

1. examine news reporting strategies of the 2020 Sabah Election Day in Malaysia and
2. explore the use of collocations between two different Malaysian newspaper portals on the 2020 Sabah Election Day.

LITERATURE REVIEW

There are a number of research on the 2020 Sabah state election, including, among others, Welsh (2021), Puyok (2021), and Wong (2020). Welsh (2021) examined how Sabah's voters shaped the election outcome and found that geography, gender, and class are crucial components of the survival strategies in voting. Puyok (2021) looked at how much pre-election voter sentiment affected the electoral outcomes, while Wong (2020) investigated the outcomes of the Sabah. Although various methodologies and data types were used, studies related to the 2020 Sabah state election have not been explored linguistically.

CORPUS-ASSISTED DISCOURSE STUDY (CADS)

The benefits of combining the CL approach with critical discourse analysis (henceforth, CDA) can be traced as far back as Baker et al., (2008). Baker et al. (2008), in their notable research on British news articles about refugees, asylum seekers, immigrants and migrants (referred to as RASIM), promoted the adoption of CL and CDA – particularly how processes such as collocation and concordance analysis were used to categorise the representation of RASIM as well as directing discourse analysts to representative texts in order to carry out qualitative analysis.

This resonates with Partington's description of a particular type of research, termed Corpus-assisted Discourse Studies (hereafter, CADS), where "the investigation and comparison of features of particular discourse types, integrating into analysis, where appropriate, techniques and tools developed within corpus linguistics" (2010: p. 88) have become a mature and popular approach in the study of text and talk (Mautner, 2019). According to Jaworska (2016), the CADS approach incorporates quantitative tools from CL while expanding the methodological paradigm through the integration of techniques typically associated with qualitative discourse analysis. This allows for a comprehensive understanding of the discourse in question and its contextual nuances. Additionally, Partington (2013) noted that CADS researchers are particularly interested in prioritising a comparative approach so as to 'compare how the phenomenon behaves elsewhere' (p.12). Hence, it is interesting to note that the CADS approach has been employed to study a variety of discourses, such as the discourse construction of 'women and work' in Italian and German texts collected from the Internet (Nardone, 2018); language ideologies in English and French-Canadian newspapers (Freake, 2012), presidential speeches (Hamed, 2021) and the representations of climate change and global warming in a newspaper (Liu & Huang, 2022).

From a cross-linguistic CADS perspective, Nardone (2018) investigated how the discourse on 'women and work' is constructed in Germany and Italy through the analysis of two general corpora from Sketch Engine, namely itTenTen and deTenTen. It was found that 'equal' (*pari* in Italian and *gleich* in German) and 'family' (*famiglia* in Italian and *Familie* in German) are two of the most frequent collocates in the analysis of women and work in the Italian corpus and in the German one, respectively. From the analysis of these collocates, more claims are deduced, such as findings that highlighted the relationship between family, women and work as the shared discourse in both corpora. Another cross-linguistic CADS study is Freake (2012), where she examines language ideologies through the use of bilingual corpora of English and French articles in Canadian newspapers. Results showed that both English and French-Canadian newspapers discuss language endangerment and that repeated collocation showed the hyper-linguistic awareness of the French language in Canada, for example, through references to the French nature of Quebec's national holiday.

Meanwhile, although not explicitly stating the CADS approach, Hamed (2021) champions that "[c]orpus – assisted analyses are recommended in the analyses of presidential discourses" (p. 150), and this could be because of the empirical nature of corpus linguistic methods that "present results that have statistical significance" (ibid). Hamed (2021) employed a corpus-assisted analysis of keywords and their collocations in the US Presidential Statements of the last four US presidents (Clinton, Bush, Obama, and Trump), and after conducting the keyword analysis, Hamed found that the US Presidential Statements or speeches reflected the Democrats and Republicans style of arguments. Similarly, Liu and Huang (2022) employed corpus-assisted analysis of keywords and their collocations to study the representations of the terms 'climate change' and 'global warming' in *The New York Times* (2019). They follow Baker et al. (2008) when they frame or give patterns to specific text by way of examining strong collocates. The examination of the top 100 lexical collocates revealed shared collocates between the two selected terms (i.e., climate change and global warming), suggesting that they are represented in a much similar way in the newspaper. In spite of the unclear labelling of CADS as their methodologies, both Hamed (2021) and Liu and Huang (2022) provided close readings of the concordance lines of collocates or made inferences when it came to analysing collocations that would reflect the real and extended instances of social interaction, that is discourse. Investigating words that are likely to appear together in discourse, as Firth (1957) rightfully notes, enables one to "[…] know a word by the company it keeps" or, put

simply: how one uses a particular word or phrase in discourse.

For the most part, the available literature on elections in Malaysia has mainly focused on the Malaysian General Elections (Funston, 2018; Mohamad Jamil, 2018; and Nadzri, 2018), not the state-level elections. These studies, together with the few which specifically examined the Sabah 2020 election (Puyok, 2021; Welsh, 2021; and Wong, 2020), predominantly focused on electoral politics and society (not linguistically). Furthermore, there have also been limited attempts to study collocations in the local political discourse – exceptions include Jalli et al., (2022) on ICERD-related news reports and Jamaluddin Aziz (2019) on broadcasting trends and issues in Malaysian Hansard corpus.

METHODOLOGY

The corpus consists of online news articles extracted from two online local newspapers, namely The Star, which is a mainstream newspaper, and Malaysiakini, an alternative newspaper, by using the search terms 'Sabah' and 'COVID-19' where focusing on only two search terms would result in a larger return of articles. It is important to note here that The Star and Malaysiakini are referred to as the online edition of its typical broadsheets, even though online editions have been argued to be slightly different in terms of style of reporting (Rogers, 2019). The selected articles correspond to the period of 30 days before the Sabah election day, during the Sabah election (26 September 2020), and 30 days after the Sabah election day ended. This period is interesting since the Sabah election happened during the COVID-19 pandemic in Malaysia, and it was reported that 70% of positive cases registered from 26 September until 12 October 2022 originated from Sabah (Batumalai, 2020). This was mainly due to the people infected whilst travelling back and forth to the Peninsular after the election, as well as mass gatherings during election campaigns by politicians who did not observe the standard operating procedures (Latiff, 2020). As far as representation is concerned, both online news portals are regarded as widely read and trusted by Malaysians - in the 2020 Reuters Institute Digital News Report,[1] Malaysiakini ranked number one for the highest weekly reach (45%) of the population referring to this source on a weekly basis (almost half of that percentage was reported to view this source more than three days per week). Meanwhile, The Star ranked third with 30% of the 81% internet penetration in Malaysia.[2] The corpus is presented as follows:

TABLE 1. Description of the Sabah Election Newspaper Articles

Corpus	Number of texts	Number of running words
The Star	422	141,178
Malaysiakini	574	249,690

---

[1] https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf
[2] In terms of online newspaper articles versus the typical broadsheets, Reuters Institute Digital News (2020) further reports that online sources of news still maintained a high percentage of preference among users (86%) since 2017.

In terms of the analytical data procedures, in corpus linguistics, classic corpus techniques like keywords and collocational analyses are frequently used in assisting discourse studies (also see Partington, 2018 on Corpus-Assisted Discourse Studies or CADS). As described earlier, CADS is beneficial to use here since we began the investigation using corpus techniques that are more quantitative in nature (e.g., keywords list) followed by close inspection of the discourse (which is more qualitative) through reading concordance lines of keywords and selected collocates (e.g., KWIC or Key Word In Context). Usually, collocational analysis is carried out after a keywords list is first automatically generated. Using #LancsBox6.0, the present study is able to conduct a keywords analysis by combining two separate word lists, as illustrated in Figure 1 below:

FIGURE 1. Illustration of how keywords analysis is produced from the 'Words' function
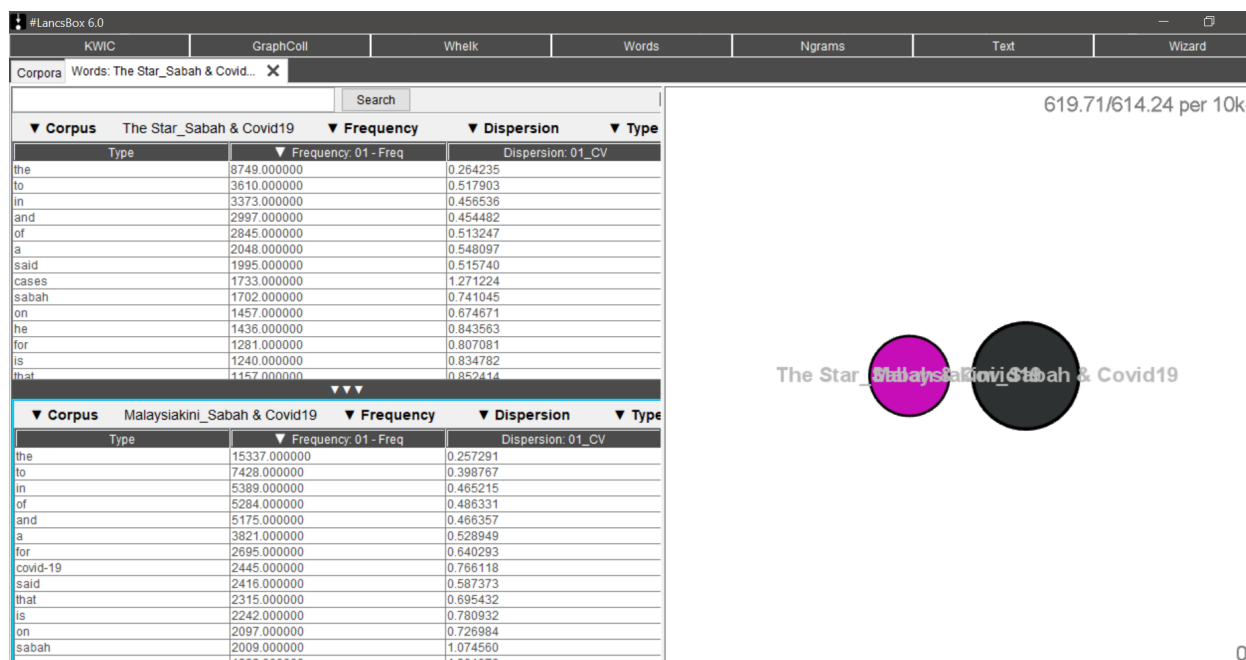

Figure 1 shows the two separate word lists generated by the software through the 'Words' function. Upon receiving the two different colour-sized circles, these circles are merely brought close together (where they overlap), and the software automatically generates a keywords list (Table 4). This is particularly useful for corpora that are not equal in size. In terms of the parameters, a minimum of 5 was selected as the cut-off point following Culpeper (2009: p. 34); where he finds this considerably sufficient given the relatively small data set. The significance test for measuring keywords was decided for log-likelihood with a probability value of smaller than or equal to 0.01. This, as Rayson (2003) describes, would be a reliable general measure for data sets that are either low in frequency, with different sizes and so on. As described earlier, following the keywords list, a collocation analysis is then performed focusing on the keyword 'total' that was significantly more frequent in The Star compared to its occurrence in Malaysiakini (see Table 4 below). More in-depth analysis is then carried out by close reading of the concordance lines. This is further discussed in the next section.

RESULTS AND DISCUSSION

HOW WAS SABAH ELECTION REPORTED IN THE ONLINE NEWSPAPERS

As previously mentioned, keywords are an important starting point in a typical corpus study. Generally, 'keywords' are defined as statistically significant lexical items (Culpeper, 2009) either frequently or infrequently occurring in one set of texts compared to their occurrence in a relative set of texts. In this paper, keywords are calculated in terms of frequency and/or infrequently occurring words in The Star compared to their occurrence in Malaysiakini. Culpeper (2009: p. 32) argued that keywords are "words whose frequencies differ significantly from their frequencies in a norm", including repetitions of these words "that statistically deviates from the pattern formed by that item in another context" (p. 4). In essence, keywords are normally referred to as quantitative measures where the sum of a target word is deemed key or salient when it is calculated in terms of a statistical measure like log-likelihood, for instance. To reiterate, keywords in this research were calculated using the log-likelihood measure with parameters as indicated above (with a probability value of smaller than or equal to 0.01 and a cut-off point maintained at 5). This, similar to Culpeper (2009), would arrive at the following:

(1) a sufficient number of meeting one's research goals,
(2) a not overwhelming number of words to analyse,
(3) an adequate dispersion of the least some keyword instances, and
(4) any one-off or extremely rare word types being minimised. Of course, whilst the previous sentence identifies important factors, there is little clarity regarding what counts as 'sufficient', 'overwhelming', 'adequate' or 'minimised'
(p. 34)

In terms of "aboutness", investigating the keywords would first reflect what the two corpora are about. Table 2 presents positive and negative keywords from The Star against Malaysiakini.[3] Positive keywords describe the more statistically significant keywords appearing in The Star compared to their occurrence in Malaysiakini, while negative keywords simply mean the opposite (*caused*, *thus*, *calls* and so on appear statistically more infrequent in The Star than in Malaysiakini). It can be seen that positive keywords in The Star mainly include content words that refer to the name Tan Sri Dr Noor Hisham Abdullah ('Datuk', 'Dr', 'Tan', 'Sri'/ 'Seri', 'Hisham', 'Noor'), 'Barisan' in reference to political party Barisan Nasional, the month 'September', and 'total' that could be used as an adjective ('total' + N). This means that The Star makes more mention of the Director General of Health Malaysia compared to Malaysiakini. Although not surprising since Dr Noor Hisham appears on the news more significantly during the pandemic in terms of COVID-19 and vaccinations, it is interesting to find that the overuse of this type of reporting can be seen in The Star compared to in Malaysiakini.

In contrast, functional words such as *caused*, *thus*, *remand*, and *traced* appear less frequently in The Star, along with lexical words like *laboratories*, *training*, *water* and *employers*. Two of the negative keywords are polysemous (*calls* and *message*) and, therefore, can be regarded as both functioning lexical and grammatical. These words are more generic in that they may refer to issues related to science (*laboratories*), environment (*water*), services (*training*, *employers*) and acts that could be associated with them (*remand*, *traced*) and, in turn, represented as more frequently occurring in Malaysiakini relative to their occurrence in The Star.

---

[3] Positive and negative keywords are determined by the software where this was generated automatically (refer to #LancsBox6.0 manual p.33 available from their website: Vaclav Brezina, Lancaster University, 2021: http://corpora.lancs.ac.uk/lancsbox/help.php)

TABLE 2. The Star keywords list[4]

	Positive Keywords	Dispersion	Statistic	Negative Keywords	Dispersion	Statistic
1	Datuk	1.54	860.12	caused	4.70	6.63
2	dr	0.92	797.93	thus	4.47	6.63
3	tan	0.86	665.34	calls	5.45	6.64
4	Sri	0.84	509.76	laboratories	8.15	6.64
5	Hisham	1.11	349.05	training	8.62	6.64
6	Seri	1.77	348.85	message	6.40	6.66
7	Barisan	2.61	325.61	remand	7.92	6.66
8	noor	1.06	321.03	traced	6.02	6.66
9	sept	1.27	253.79	water	9.58	6.66
10	total	1.49	240.86	employers	9.78	6.91

Since 'total' was found to be among the top positive keywords in The Star (i.e., 'total' occurred less significantly in Malaysiakini and can be seen in Appendix A) after words that merely refer to the Director General of Health, it would be interesting to examine how this word is used similarly or differently in the two corpora. This was further analysed in terms of a collocation analysis using GraphColl, which will be presented next.

NEWS REPORTS RELATED TO SABAH ELECTION: A FOCUS ON COLLOCATES OF 'TOTAL'

This section begins by presenting the collocates of 'total' in The Star, followed by 'total' in Malaysiakini ('total' was identified as a target word to look at based on the analysis of the keywords that makes it more empirical and corpus-driven in nature). As a rule of thumb, the criteria set for the identification of collocates via #LancsBox6.0 was between a span of 5 words to the right and left, using T-score as the statistical measure[5] (increasing the default value to 6, where we were able to work with more restricted collocates) and collocation frequency was set at a minimum of 5. This means that in order for collocates to be selected, they must appear more than five times across texts in the corpora. Figures 2 and 3 present collocates of the keyword 'total' in The Star and Malaysiakini, respectively. In The Star, 'total' co-occurred with 17 collocates, as can be seen in Figure 2, ranging from statistically frequent functional words like *of* and *the* (seen closer to the word 'total' representing the high statistical count: 19.94; 18.37), followed by lexical words like 'cases' that was ranked third (16.42), followed by *number* (12.87), *a* (12.28), *in* (10.24), *to* (9.67), *bringing* (8.56), *patients* (8.22), *active* (7.4), *new* (7.39), *or* (7.67), *recoveries* (7.11), *since* (6.71), *covid-19* (6.49), *were* (6.34) and *country's* (6.19). On the other hand, 'total' is collocated only

---

[4] Keywords refer to the salient words in a corpus whose frequency is unusually high (positive keywords) or low (negative keywords) in comparison with a reference corpus.

[5] A similar analysis using MI-value (with the same restrictions: threshold value increased to 6 instead of the standard three while collocation frequency remained at the default minimum of 5 occurrences) resulted in lower frequency collocates such as words like 'bring', 'bringing' and 'cumulatively' that ultimately refers to the collocation between 'total' + 'of' with regard to 'cases'.

frequently with six words in Malaysiakini, and they are *of* (14.78), *a*n (11.64), *the* (11.55), *cases* (10.39), *in* (7.78) and *to* (7.34) as shown in Figure 3.
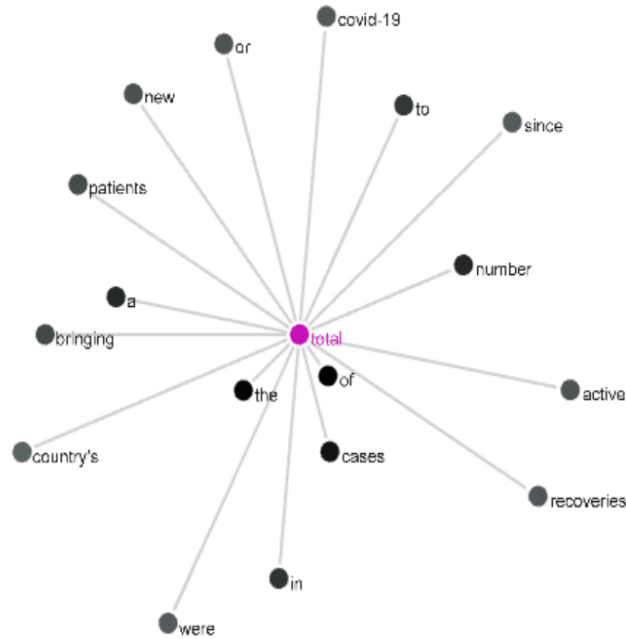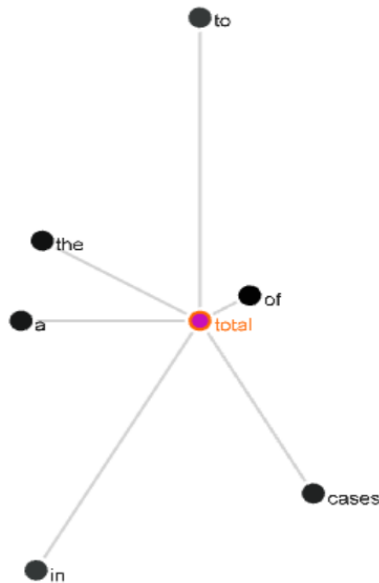

FIGURE 2. Collocates of 'total' in The Star

FIGURE 3. Collocates of 'total' in Malaysiakini

Since both corpora showed statistically more frequent collocations between the words 'total' and 'of', further collocation analysis was carried out to see whether these words shared any similar collocates. This was done by simply double-clicking on the word 'of' in the GraphColl image, and as a result, it was found that 'cases' was seen to be a commonly shared (lexical) collocate of 'total' and 'of' in both The Star (1,733 times) and Malaysiakini (1,993 times) as can be seen in Tables 3 and 4 below. Further double-clicking on the shared collocate 'cases' reveal the collocation network presented in The Star (see Figure 4) and Malaysiakini (see Figure 5) that further demonstrate how these words are used differently in discourse and, again, considered to be more empirical than mere preconceived assumptions.

TABLE 3. Shared collocates between 'total' + 'of' in The Star

	Collocate	Frequency
1.	the	8749
2.	to	3610
3.	in	3373
4.	a	2048
5.	**cases**	**1733**
6.	covid-19	1086
7.	were	904
8.	new	841
9.	number	374
10.	or	267
11.	patients	209
12.	active	197
13.	since	154
14.	bringing	113
15.	country's	105
16.	recoveries	64

TABLE 4. Shared collocates between 'total' + 'of' in Malaysiakini

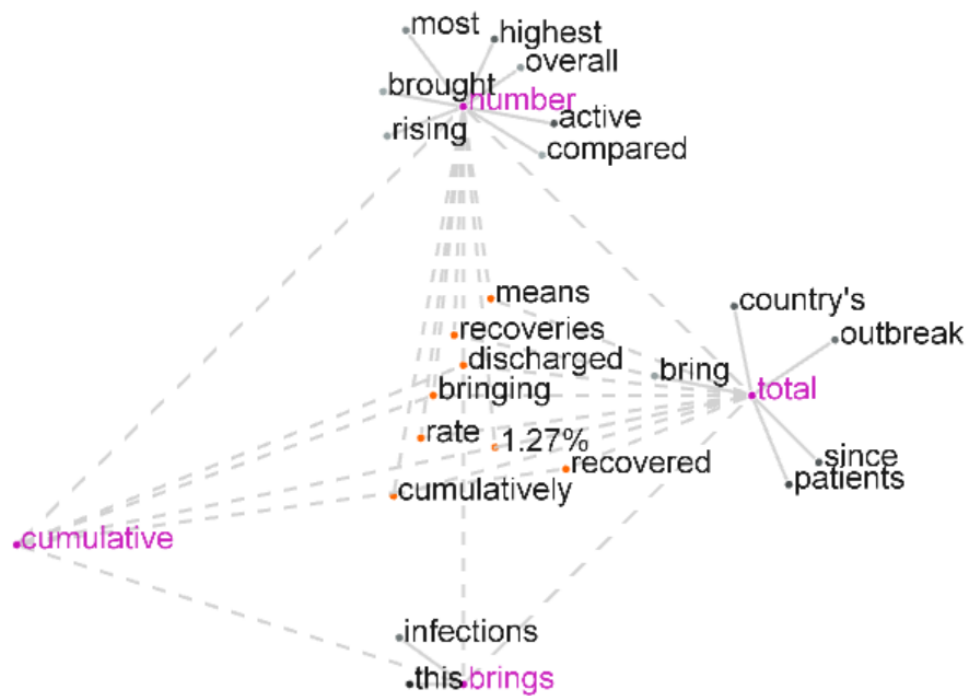|	Collocate	Frequency
1.	the	15337
2.	to	7428
3.	in	5389
4.	a	3821
5.	**cases**	**1993**

FIGURE 4. Collocation network for 'total' in The Star using MI-score

FIGURE 5. Collocation network for 'total' in Malaysiakini using MI-score

Briefly, collocation networks in Figures 4 and 5 above show how connections between the highlighted words were made in the two local newspapers. More specifically, 'total' was used in The Star with 'brings', 'number', and 'cumulative' relating to 'recoveries', whereas 'total' in Malaysiakini co-occurred with the same collocates in terms of being 'active'. Another important thing to consider is that while The Star makes mention of (total) 'infections' that are signalled by the collocate 'brings', in Malaysiakini, this connection is strongly associated with 'death'. However, in order to attribute any of the collocations above to the 2020 Sabah election, the next part of the paper will examine concordance lines for 'cases' by specifying lines that also include 'Sabah' in order to focus on frequently occurring collocations in relation to the 2020 Sabah election. This would ensure that no sweeping generalisations are made.

FUNCTIONAL ANALYSIS OF SHARED COLLOCATES IN BOTH CORPORA:
'TOTAL', 'OF' AND 'CASES'

To reiterate, both corpora were collected based on reports that were obtained from keying in the two search terms: 'Sabah' and 'COVID-19'. This allowed for reports from both media agencies to be mutually representative of the context and setting. Since the frequent collocation 'total' + 'of' was found to be occurring in both The Star and Malaysiakini with 'cases', this warrants further investigation as to how these occurrences or patterns functioned in the newspapers relating to COVID-19, particularly the Sabah election (as a reminder, concordance lines were specified to those including 'Sabah' only). In so doing, concordance lines for these collocations in the two corpora were examined closely, and collocations that did not refer to COVID-19 and Sabah

election were discarded. This resulted in 64 lines for 'Sabah' co-occurring with 'of' + 'cases', and three lines where 'Sabah' collocates with 'of' and 'total' in Malaysiakini; 37 lines for 'Sabah' co-occurring with 'of' + 'cases', and four lines with 'of' and 'total' in The Star as shown in Figures 6, 7, 8 and 9 respectively. They are discussed in terms of three recurring themes, namely, Sabah post-state election's direct effect, the election's spill-over effects, and assumptions/speculations.[6]

SABAH POST-STATE ELECTION'S DIRECT EFFECT

For the most part, concordance lines of the collocation 'of' + 'cases' + 'Sabah' in both corpora similarly reflect reporting of Sabah post-state election's direct effect on the number of COVID-19 cases in Malaysia at the time. Figures 6 and 8 present concordance lines for 'of' occurring with 'cases' and 'Sabah' in The Star and Malaysiakini, respectively. While there was more mention of this collocation in Malaysiakini (15 lines compared to 5 lines in The Star), it can be generally seen that the number of total of cases referring to COVID-19 involved the mentioning of Sabah (e.g., '*involving travellers returning* from Sabah', '*recently travelled to* Sabah', '*with travel history to* Sabah' and others). The Star also showed similar reports through the use of phrases like 'x number of the [...] cases *were from Sabah/ detected in Sabah/ recorded in Sabah*', as shown in Figure 6.

However, it can be argued that The Star displayed more direct expressions of Sabah in contrast to Malaysiakini reports that are rather hedged, for instance: 'To come *from Sabah*', 'continued *to be from Sabah*', 'linked to/detected in/recorded *in Sabah*' as opposed to 'involving travellers returning from/ recently travelled to/ with travel history to Sabah, and others (see Figure 7). In addition, there were also numerous uses of quantifiers/evaluative markers like 'positive', 'high', 'highest/large/alarming rise/rising number of [the] cases' referring to Sabah in The Star compared to Malaysiakini. These initial findings (despite the style of reporting) imply that the Sabah 2020 state election was deemed mainly contributing to the COVID-19 pandemic, given its assumed description as a mass gathering event. As Herng et al. (2022) found, "the National Daily estimated Rt increased to 1.72 during the Sabah electoral process, which was a mass gathering event" (p. 11). In fact, Malaysiakini reported that due to the Sabah electoral process, cases tripled as "doubling time for Sabah [...] swiftly changed as the campaign started for the Sabah election" (Malaysiakini, 16 October 2020).

1	SEPT_29_4_Covid_The Star.txt	Sri Dr Noor Hisham Abdullah said 73	**of**	the *cases* were from *Sabah*, Selangor (14),
2	OCT_09_3_Covid-19_ The_Star.txt	Sri Dr Noor Hisham Abdullah said 274	**of**	the *cases* were detected in *Sabah*, while
3	OCT_18_2_SOP_The_St ar.txt	a total of 224 *cases* or 25.8%	**of**	the day's total cases. "Comparatively, for *Sabah*,
4	OCT_19_2_ Total_cases The_Star.txt	and 70. He noted that 702 or 80.6%	**of**	the new *cases* were detected in *Sabah*,
5	OCT_26_2_ cases_The_Star.txt	been a reduction of 356 *cases* (40%	**of**	total cases) recorded in *Sabah* compared with

FIGURE 6. Specific Number Before Collocation 'of' in The Star

---

[6] The first two themes were taken from Lim et al. (2021) where they reported that: an estimated (i) 70.0% of COVID-19 case counts within Sabah post-state election were attributable to the election's direct effect; (ii) 64.4% of COVID-19 cases in the rest of Malaysia post-state election were attributable to the election's spill-over effects.

6	SEPT_30_7_linked_to_Sabah_The_Star.txt	Malaysia have recorded new Covid-19 cases *because*	**of**	individuals *who returned from Sabah*. On Wednesday
7	SEPT_30_13_New_The_Star.txt	Malaysia have recorded new Covid-19 cases *because*	**of**	individuals *who returned from Sabah*. On Wednesday
8	OCT_11_6__Labuan_says_The_Star.txt	and Saudi Arabia (one). The highest number	**of**	new cases *were recorded in Sabah* (488)
9	OCT_12_3_Five_new_clusters_The_Star.txt	and Saudi Arabia (one each). The highest number	**of**	localised cases were recorded in Sabah with
10	OCT_14_2__deaths_The_Star.txt	Dr Noor Hisham Abdullah said the bulk	**of**	the *cases continued to be from Sabah,*
11	OCT_18_4_Sabah__DG_The_Star.txt	in the number of cases is *because*	**of**	the high number of *samples from Sabah*
12	OCT_19_2_20000_The_Star.txt	in the number of cases is *because*	**of**	the high number of *samples from Sabah*
13	OCT_23_3_death_toll in The_Star.txt	of them requiring ventilator support. The majority	**of**	*the cases continued to come from Sabah,*
14	OCT_24_1_single_day_recorded_The_Star.txt	28 of them requiring ventilator support. The majority	**of**	*the cases continued to come from Sabah,*
15	OCT_24_4_plus_seven_deaths_The_Star.txt	on Saturday (Oct 24), the highest number	**of**	cases to date, with *Sabah contributing 889*

FIGURE 7. The direct effect of Sabah on the Covid-19 cases in The Star

16	SEPT_26_4_TRAVEL IN SABAH_MK.txt	today's cases, Noor Hisham (photo) said 33	of	the 64 *cases* in Sabah involve the
17	OCT_01_10_RESTRICT SABAH_MK.txt	to noon today. A total of 118	of	these *cases* were in Sabah; 98 in
18	OCT_09_7_TO CONTAIN_MK.txt	through Facebook. There have been 3,148 *cases*	of	Covid-19 in Sabah since the first cluster
19	OCT_09_6_DEATHS IN SABAH_MK.txt	Health Ministry has reported 354 new cases	of	Covid-19 and six deaths, all in Sabah,
20	OCT_11_2_HIGHEST ON RECORD_MK.txt	comprise 494 Malaysians and 59 non-citizens. 11	of	these *cases* are peninsular returnees from Sabah.
21	OCT_17_3_38 DEATHS OVER 10 DAYS_MK.txt	Covid-19: New record high	of	869 *cases*, 38 deaths in Sabah over
22	OCT_18_PATIENTS MADE TO WAITMK.txt	871 new cases recorded today, with 702	of	the *cases* coming from Sabah alone.
23	OCT_26_8_1240 NEW CASES_MK.txt	Health Ministry reported 1,240 new Covid-19 *cases,*	of	which 927 were reported in Sabah alone.

FIGURE 8. The specific number before collocation 'of' in Malaysiakini

24	SEPT_28_3_QUARAN TINED TOO_MK.txt	*view of the surge in the number*	**of**	cases in Sabah and will conduct important
25	SEPT_29_1_HIGH RISK TRAVEL_MK.txt	Health Ministry. *This came following a surge*	**of**	Covid-19 cases in Sabah and increased travel
26	OCT_01_8_ANOTHER 4 CLUSTERS_MK.txt	by the relevant district health offices. Overall,	**of**	the 118 cases in Sabah today, 74
27	OCT_01_10_FROM SABAH_MK.txt	hours up to noon today. A total	**of**	*118 of these cases were in Sabah;*
28	OCT_07_5_NEW CLUSTER_MK.txt	acquired infections and two are imported. Most	**of**	the *new cases are reported in Sabah*
29	OCT_09_6_SIX DEATHS_MK.txt	looks to be *on the rise*. Most	**of**	the *cases were in Sabah*. Is this
30	OCT_12_9_MCO SABAH_MK.txt	cases of Covid-19- a new record. Most	**of**	the *active cases are in Sabah* (2,783).
31	OCT_13_6_TRAVEL QUARANTINE_MK.txt	and considering the *spike in the number*	**of**	Covid-19 cases in Sabah, all travellers arriving
32	OCT_16_1_REFLECT SEVERITY_MK.txt	Awang Bulgiba Awang Mahmud (above). "The number	**of**	*critical cases is rising rapidly in Sabah*
33	OCT_16_4_BREACH 6000_MK.txt	in Noor Hisham said the *high number*	**of**	*new cases from Sabah* had to do
34	OCT_17_2_SABAH ICUS_MK.txt	said. Malaysia recorded 869 new cases today,	**of**	*which 451 came from Sabah.*
35	OCT_18_1 ELECTION SAYS CM_MK.txt	the next six months given *the surge*	**of**	Covid-19 *cases in Sabah after it held*
36	OCT_25_3_NEEDS AID_MK.txt	isolation of infected individuals. *Yesterday's record-high number*	**of**	*new cases* (1,228) saw 889 in Sabah

FIGURE 9. The direct effect of Sabah on the Covid-19 cases in Malaysiakini

Closer inspection reveals that people were cautioning against the state electoral process as there were reports stating the likelihood or tendencies for politicians to be tested positive upon returning from Sabah. This can be seen in the highlighted text below:

> COVID-19| Media advocacy group Gerakan Media Merdeka (Geramm) has urged **politicians who are being quarantined after campaigning in Sabah to reveal whether they had tested positive or negative for Covid-19.** This came after New Straits Times Press Berhad said one of its employees who returned from covering the Sabah state election subsequently tested positive for the coronavirus. Geramm said it is aware of concerns raised by reporters amid the rising number of Covid-19 cases and urged for cooperation from all parties to be aware of their health status in addition to practising self-restraint to prevent the spread of the disease. This is especially for national leaders inviting the press to provide coverage. "It cannot be denied that the **recent busyness of national leaders has indirectly increased the risk of infection, especially if there is a history of travel to districts that have recorded a high number of cases.** "Political leaders returning from Sabah and undergoing quarantine are urged to update their health status, whether they are negative or positive for Covid-19," said the group in a statement today. (*Malaysiakini*, 5 October 2020)

THE ELECTION'S SPILL-OVER EFFECTS

Another finding from concordance lines showed that Malaysiakini reported Sabah to be one of the "high-risk areas" since COVID-19 cases were reported coming from the state (particularly based on nine screening of arrivals reported between 27 September and 4 October). This implied plausible causes for infections to be on the rise, especially after the state election was held. Meanwhile, there were also examples – corroborated past research (Lim et al., 2021) that there was a spill-over from the Sabah state elections to other states in Malaysia – indicated by 'of cases' co-occurring with 'linked to' or 'tied to' such as in lines 1-5 in Figure 10 below. Lim et al. (2021)

argued that events that took place during the Sabah electoral process "demonstrate[s] both the direct risk of COVID-19 transmission from mass gatherings and the potential for these mass gatherings to spill over to other regions, resulting in an outbreak of national scale" (p. 6).

1	SEPT_30_linked_to_Sabah_TheStar.txt	Abdullah *said Selangor* had the most number	**of**	*cases linked to Sabah* with five, followed
2	SEPT_30_7_cases_detected_TheStar.txt	Abdullah said *Selangor* had the most number	**of**	*cases linked to Sabah* with five, followed
3	OCT_01_3_tied_to_Sabah_TheStar.txt	*Selangor* has most number	**of**	cases *tied to Sabah* returnees Thursday, 01
4	OCT_01_3__Sabah_returnees_TheStar.txt	(pic) said *Selangor* had the most number	**of**	*cases linked to Sabah* with five cases,
5	OCT_03_1_Highest_daily_TheStar.txt	visiting Sabah recently. "This makes the number	**of**	*positive cases linked to Sabah travel* since

FIGURE 10. Spill-over from Sabah state election to other states in The Star

6	SEPT_27_CASES RISE TO 950_MK.txt	cases) Setapak cluster (new) 2 cases Screening	**of**	arrivals from high-risk areas in Sabah: 1
7	SEPT_27_4_150 NEW COVID19_MK.txt	1 case Selangor (10 new cases) Screening	**of**	arrivals from high-risk areas in Sabah: 8
8	SEPT_27_4_ACTIVE CASES RISE_MK.txt	1 case Johor (2 new cases) Screening	**of**	arrivals from high-risk areas in Sabah: 2
9	SEPT_27_4_150 NEW COVID19_MK.txt	2 cases Pahang (4 new cases) Screening	**of**	arrivals from high-risk areas in Sabah: 3
10	SEPT_27_2_TRACED TO SABAH_MK.txt	Of the 15 cases detected through screening	**of**	people returning from high-risk areas Sabah, eight
11	SEPT_29_1 BEYOND SABAH_MK.txt	COVID-19| About half of Covid-19 cases outside	**of**	Sabah has been linked to the outbreak
12	SEPT_29_3_SABAH RETURNEES_MK.txt	About half of the Covid-19 cases outside	**of**	Sabah have been linked to the outbreak
13	SEPT_30_2_SABAH CASES_MK.txt	day before. Since that date, 99 cases	**of**	local transmissions have been linked to Sabah,
14	SEPT_30_2_REPORTED NATION_MK.txt	percent of locally transmitted Covid-19 cases outside	**of**	Sabah since Sept 23 is linked to
15	SEPT_30_LAX QUARANTIN_MK.txt	to 16. "Just looking at the number	**of**	cases today linked to Sabah, perhaps we
16	SEPT_30_4_MOST STATES_MK.txt	to Sabah. That brings the total number	**of**	cases involving travellers returning from Sabah to
17	OCT_02_13_SABAH TOP LIST_MK.txt	recently travelled to Sabah. The total number	**of**	cases involving recent travels to Sabah stood
18	OCT_03_2_HELP THEM_MK.txt	by a third wave of cases, many	**of**	which were traced back to Sabah where
19	OCT_04_4_SABAH WORST HIT_MK.txt	cluster: One case Selangor (31 cases) Screening	**of**	arrivals from high-risk areas in Sabah: 13
20	OCT_04_293 CASES SABAH_MK.txt	screening: Two cases Johor (Eight cases) Screening	**of**	arrivals from high-risk areas in Sabah: Six

FIGURE 11. Spill-over from Sabah state election to other states in Malaysiakini (cont. in Appendix B)

As noted, in Sabah, interstate travel was allowed during the election period, with in-person voting mandated within each individual's registered district: this led to an influx of around 250,000 Sabahan voters returning from other parts of Malaysia including politicians and campaign workers flying in from Peninsular Malaysia for the same purpose (Lim et al., 2021). This, according to Lim

et al. (2021), infers that some of the interstate travellers contracted COVID-19 during the election period due to possible mixing and "that these individuals may have played roles in seeding new outbreaks in other states upon their returns from Sabah" (p. 7). Their results showed a relationship between the substantial increase in case counts in other regions of Malaysia with the Sabah state election, which transpired in the concordance lines that can be seen in Figure 11. In a separate report, Anuar (2020) states that "[t]he Prime Minister also blames Sabah election as among the causes of huge infection surge, [mainly] linked to the movement of politicians and voters between Sabah and Peninsular Malaysia" (p. 280).

> PUTRAJAYA: Nearly all of the states in Malaysia have recorded new **Covid-19 cases because of individuals who returned from Sabah**. Health director-general Tan Sri Dr Noor Hisham Abdullah (pic) said Selangor had the most **number of cases linked to Sabah** with five cases, followed by Sarawak, Terengganu and Kuala Lumpur– which each state having three cases related to travel to Sabah.
>
> (*The Star*, 30 September 2020)

> PETALING JAYA: **Sabah has recorded more than one-third of all Covid-19 cases in the country**, as the number of cases in Malaysia breached the 20,000 mark on Sunday (18 October). Health director-general Tan Sri Dr Noor Hisham Abdullah said a total of 7,439 cumulative cases, or 36.3% of the country's overall cases, have been recorded in Sabah, with 702 new cases reported on Sunday. "The surge in the number of cases is because of the high **number of samples from Sabah** that were sent to the Institute for Medical Research (IMR)," he said during a daily press briefing on the Covid-19 situation.) He said there was a backlog in the number of tests, but it has been cleared. Dr Noor Hisham said that the ministry had conducted a total of 251,030 Covid-19 tests on individuals in Sabah, with the **infectivity rate at 2.96%**.)
>
> (*The Star,* 18 October 2020)

ASSUMPTIONS/SPECULATIONS

So far, we have seen collocations 'of' + 'cases' + 'Sabah' in The Star and Malaysiakini that have highlighted the state election's effects (both direct and spill-over) on the number of COVID-19 in the country at the time. We argued then: to what extent are these reports speculative, and are there any differences between how the events were reported between the two online news portals? Below is an excerpt from concordance lines that revealed possible scepticism coming from several sources. This, in particular, shows how persons in authority were aware of the situation but understated the implications it (the state electoral process) brought.

> As for the alarming rise of cases in Sabah after the 26 September state elections, Prof Sazaly said the real issue was the actual infection rate in the community even before the elections. "Most likely, it has been in the community even before the elections. **On the increasing cases of Covid-19 linked to the Sabah state election, the Santubong MP did not agree for a health emergency to be declared with the purpose of not holding the state election. This was because, he said, such a declaration has never been announced in the world, just to avoid facing the constituents.** "Covid-19 can be handled with everybody complying with SOPs such as physical distancing, wearing face masks, and avoiding crowded places. The matter (health) cannot be considered as an emergency," he said.   (*The Star,* 13 October 2020)

To examine these further, similar concordance steps were carried out to closely look at the collocation between 'total' + 'cases' + 'Sabah' in The Star and Malaysiakini (as a reminder, the keyword 'total' was found frequent in both corpora).

1	to 163 cases or 1.0% of the **total** *cases*. All four deaths are in *Sabah*
2	to 167 cases, or 1.0% of the **total** *cases*. All four deaths are in *Sabah*:
3	224 cases or 25.8% of the day's **total** *cases*. "Comparatively, for *Sabah*, although recording the
4	a reduction of 356 cases (40% of **total** *cases*) recorded in *Sabah* compared with Saturday.

FIGURE 12. Concordance lines for 'total' occurring with 'cases' and 'Sabah' in The Star

From Figure 12, we can see that the collocation 'total'+ 'cases' and 'Sabah' was used to refer to the total number of COVID-19 *death* cases reported in The Star, of which two of the lines indicated casualties reported in Sabah, while the remaining two lines reported a reduction in the number of cases post-state election (Herng et al., 2022: p. 7). This supported findings from reports like Lim et al. (2021) and Herng et al. (2022), where the four COVID-19 deaths were from Sabah (e.g., Sandakan, Kota Kinabalu) and the reduction of cases referred to the effects of CMCO (Conditional Movement Control Order), which was implemented after the Sabah electoral process.

1	any *cases* are Nabawan and Keningau. In **total,** there were 1,206 active *cases* in *Sabah*
2	*cases* as of noon today, bringing Malaysia's **total** active *cases* to a record 5,039. *Sabah*
3	2,023. At the time of writing, the **total** number of active *cases* in *Sabah* is

FIGURE 13. Concordance lines for 'total' occurring with 'cases' and 'Sabah' in Malaysiakini

Contrary to The Star, which reports total death cases due to COVID-19 casualties in Sabah, Malaysiakini reports 'total' collocating with 'active cases' co-occurring with a specific number of cases, as shown in Figure 13. This can be argued to be more specific in highlighting the increased number of COVID-19 cases during the Sabah electoral process and, in turn, supports Malaysiakini's assumptions of the consequences of the projected "spill-over" as discussed previously. The below example showed 'total' referring to Malaysia's total active cases, but Sabah was reported to remain the highest number of new cases, thus implying that there is a spill-over from the Sabah state election (as discussed in the previous sub-section).

> 563 new Covid-19 cases, active cases breach the 5,000 mark
> Published: 12 October 2020, 6:38 PM
>
> COVID-19| Malaysia has recorded 563 new Covid-19 cases as of noon today, bringing Malaysia's **total active cases to a record 5,039. Sabah continued to have the highest number of new cases (291)**, followed by Penang (141), where the Penjara Reman cluster accounted for all the new cases.
>
> (*Malaysiakini,* 12 October 2020)

Even when 'total' was found co-occurring with the phrase 'still uncertain' in line (3) above, it could still be argued that this implies a concern that COVID-19 cases were on the rise and is expressed in the extended excerpt taken from the report:

> At the time of writing, the **total number of active cases in Sabah is still uncertain**. At noon yesterday, the figure stood at **2,783, which was slightly more than half of the total number of active cases** in the country today.

In fact, the collocation between 'total' and 'active cases' in significant lines reported in Malaysiakini go as far as to show the highlighted adjective 'active', which functions to indicate the continuing report or rise of COVID-19 cases as a result of the Sabah electoral process. This has been shown in data reports from the Malaysian Ministry of Health (2020), for instance, one shown here in Figure 14. In his daily press release, the Director-General of Health Malaysia, Dr Noor Hisham Abdullah, was reported to have said

> "[f]or this third wave, Selangor began to see an increase in cases at the end of September 2020 after the entry and return of travellers from Sabah. The first cluster reported in Selangor in this third wave is the Jalan Apas Cluster i.e. on 28 September 2020, involving families returning from Sabah. The increase of local transmission cases in Selangor began to be seen on 2 October 2020 and have increased consistently until today"
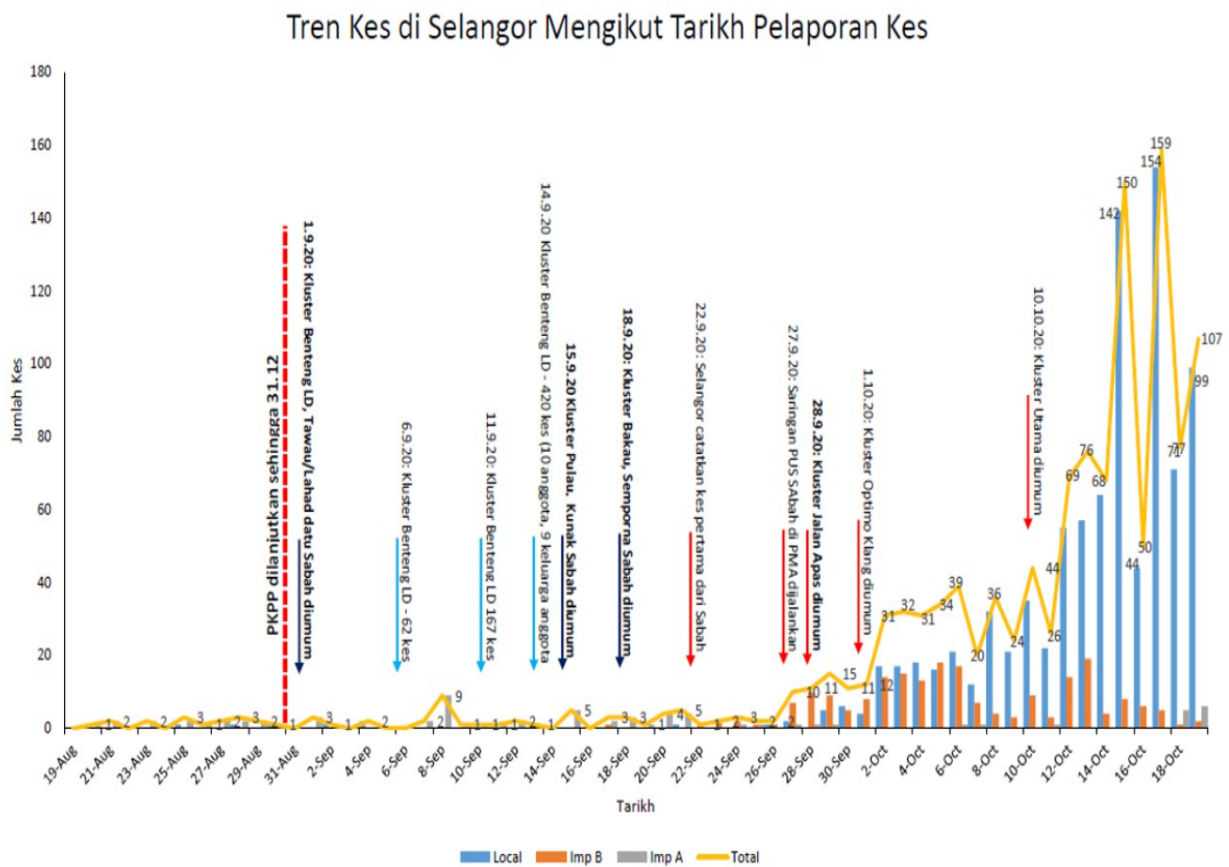>
> (Ministry of Health, 2020)

FIGURE 14. COVID-19 case trends in Selangor and steady rise post-Sabah electoral process (27.09.2020) [Source: Ministry of Health, 2020]

CONCLUSION

This case study demonstrated how media representations of the 2020 Sabah state election during COVID-19 could be explored using collocation analysis, particularly via the #LancsBox6.0 tool. Our major findings revealed that both newspapers described how total COVID-19 cases during the 2020 Sabah state electoral period cumulatively added to the existing numbers of cases in Malaysia. In particular, initial keywords analysis suggested that the statistically significant word 'total' was often used in relation to the preposition 'of' and noun 'cases', which reflected on what the news was about in general: "total cases" [of COVID-19]. More specifically, concordance lines that specified how frequent collocations occurred in terms of the Sabah state election resulted in different findings between The Star and Malaysiakini, reflecting the nature of these online newspapers as mainstream and online (alternative) media, respectively. Although both online news portals reported the significance of the Sabah state election to the number of COVID-19 cases in Malaysia, The Star made more direct reference to it by using the existential be-verb 'is' (*is from Sabah*) and spatial deictic marker 'in' (cases *in Sabah*) while also using evaluative markers like 'positive', 'high', 'highest/large/alarming rise/rising number of [the] cases' in relation to Sabah. However, Malaysiakini reported more explicitly in terms of the spill-over effects of the Sabah state election on Malaysia's overall COVID-19 cases (nine reports of screening evaluations of COVID-19 cases from Sabah and further look at the context that signalled a high infectivity rate), which also corroborated with other scientific studies surrounding the issue (e.g., Lim et al., 2021; Herng et al., 2022). Interestingly, The Star showed more "speculated" news as some of the collocations were found to be from politicians' verbatim that indicated their weak arguments on the situation during the time of the Sabah state election. This was not found in Malaysiakini and therefore presented critical reading when it comes to media reporting of sensational/controversial news. Also, the differences in the way news were reported in The Star and Malaysiakini during this timeline reflect the evident role of the media in the run-up to any election, specifically in the Malaysian COVID-19 context. Since the Internet and online news have become a medium of political news and analysis, online newspapers should strive to create a sense of credibility to make their reporting more believable.

On a general level, this paper illustrates linguistic methods of enquiry, specifically collocation analyses on news reporting of the Sabah 2020 State Election Day in Malaysia that also took place during the COVID-19 pandemic. This was done through the use of corpus linguistic methods by comparing selected news articles from The Star (mainstream newspaper) with Malaysiakini (alternative news) that were published during the 2020 Sabah election period. Beginning with a keywords analysis, underpinned by the CADS approach, this integration between analysing discourse and using corpus linguistic techniques was suitable to empirically guide our research to ascertain combinations of key phrases that have, in turn, provided us with the findings discussed above. Limitations include that due to space, analysis of the collocation network that would give us insight into second, third, and even fourth-order collocates was not possible and, therefore, can be a potential avenue for further research. Future studies on other news media reporting on the same issue are suggested to examine whether they show similar collocational patterns in the news reports. The effects of these different news media on the public perception of the issue are also worth exploring.

ACKNOWLEDGEMENTS

REFERENCE

Anuar, H. M. (2020). Election During Covid-19 Pandemic: Constitutional Perspectives. *International Journal of Law, Government and Communication*, *5*(21), 277–284. DOI: 10.35631/IJLGC.5210023.

Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyzanowski, M., McEnery, T., & Wodak, R. (2008). A Useful Methodological Synergy? Combining Critical Discourse Analysis and Corpus Linguistics to Examine Discourses of Refugees and Asylum Seekers in the UK Press. *Discourse & Society, 19*(3), 273–306.

Baker, P., Gabrielatos, C., & McEnery, T. (2013). Sketching Muslims: A corpus-driven analysis of representations around the word 'Muslim'in the British press 1998–2009. *Applied linguistics*, *34*(3), 255-278.

Batumalai, K. (2020, October 16). *Malaysia's Sporadic Covid-19 Cases Spiking, Driven By Sabah Surge.* CodeBlue. https://codeblue.galencentre.org/2020/10/16/malaysias-sporadic-covid-19-cases-spiking-driven-by-sabah-surge/

Brezina, V. (2018). Collocation Graphs and Networks: Selected Applications. In: Cantos-Gómez, P., Almela-Sánchez, M. (eds) *Lexical Collocation Analysis. Quantitative Methods in the Humanities and Social Sciences*. Springer,

Brezina, V., Weill-Tessier, P., & McEnery, A. (2021). #LancsBox v. 6.x. [software package]

Cham. https://doi.org/10.1007/978-3-319-92582-0_4

Culpeper, J. (2009). Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's *Romeo and Juliet*, *International Journal of Corpus Linguistics*, *14*(1), 29–59.

Firth, J. R. (1957). *Papers in linguistics, 1934-1951*. London: Oxford University Press.

Freake, R. (2012). A cross-linguistic corpus-assisted discourse study of language ideologies in Canadian newspapers. In *Proceedings of the Corpus Linguistics Conference (CL2011)*, pp. 1–11.

Funston, J. (2018). Malaysia's 14th general election (GE14) – the contest for the Malay electorate. *Journal of Current Southeast Asian Affairs, 37*(3), 57–83.

Hamed, D. M. (2021). Keywords and collocations in US presidential discourse since 1993: A corpus-assisted analysis. *Journal of Humanities and Applied Social Sciences*, *3*(2), 137–158.

Herng, L. C., Singh, S., Sundram, B. M., Zamri, A. S. S. M., Vei, T. C., Aris, T., Ibrahim, H., Abdullah, N. H., Dass, S. C., & Gill, B. S. (2022). The effects of super spreading events and movement control measures on the COVID-19 pandemic in Malaysia. *Scientific Reports, 12*(1). https://doi.org/10.1038/s41598-022-06341-1

Jaafar, E. A. (2022). Collocation Networks of Selected Words in Academic Writing: A Corpus-Based Study. *Arab World English Journal (AWEJ) 13*(1), 240 - 255. https://dx.doi.org/10.24093/awej/vol13no1.15

Jalli, N., Joharry, S. A., & Salleh, S. M. (2022). ICERD in Malaysian online news reports: Analysis of rhetoric and public opinion. *Social Sciences & Humanities Open*, *6*(1), 100318. https://doi.org/10.1016/j.ssaho.2022.100318

Jamaluddin, A. (2019). Retrieving trends and issues of penyiaran/broadcasting in Malaysian parliamentary debates from 1957 to 2018: a culturomics approach. - *Jurnal Komunikasi, Malaysian Journal Of Communication, 35*(4), 172-192.

Jaworska, S. (2016). Using a Corpus-Assisted Discourse Studies (CADS) Approach to Investigate Constructions of Identities in Media Reporting Surrounding Mega Sports Events: The Case of the London Olympics 2012. In I. Lamond & L. Platt (Eds.), *Critical event studies: Approaches to research* (pp. 149–174). London: Palgrave Macmillan. https://doi.org/10.1057/978-1-137-5

Joharry, S. A., & Turiman, S. (2020). Collocation Networks and Covid-19 in Letters to the Editor: A Malaysian Case Study, *Asia Pacific Journal of Corpus Research*, *1*(1), 1-30. https://doi.org/10.22925/apjcr.2020.1.1.1

Lai, H. (2019). Collocation analysis of news discourse and its ideological implications. *Pragmatics, 29*(4), 545–570. https://doi.org/10.1075/prag.17028.lai

Latiff, R. (2020). *Malaysian leaders draw flak after post-election virus jump.* REUTERS. https://www.reuters.com/article/healthcoronavirus-malaysia-idUSL4N2GT0US

Lim, J. T., Maung, K., Tan, S. T., Ong, S. E., Lim, J. M., Koo, J. R., Sun, H., Park, M., Tan, K. W., Yoong, J., Cook, A. R., & Dickens, B. S. L. (2021). Estimating direct and spill-over impacts of political elections on COVID-

19 transmission using synthetic control methods. *PLoS Computational Biology*, *17*(5): e1008959. https://doi.org/10.1371/journal.pcbi.1008959

Liu, M., & Huang, J. (2022). "Climate change" vs "global warming": A corpus-assisted discourse analysis of two popular terms in The New York Times. *Journal of World Languages, 8*(1), 34–55. https://doi.org/10.1515/jwl-2022-0004

Mautner, G. (2019). A research note on corpora and discourse: Points to ponder in research design. *Journal of Corpora and Discourse Studies*, *2*, 1–13. https://doi.org/10.18573/jcads.32

Ministry of Health. (2020). UPDATES ON THE CORONAVIRUS DISEASE 2019 (COVID-19) - SITUATION IN MALAYSIA [Press release]. https://kpkesihatan.com/2020/10/20/kenyataan-akhbar-kpk-20-oktober-2020-situasi-semasa-jangkitan-penyakit-coronavirus-2019-covid-19-di-malaysia/

Mohamad Jamil, S. N. (2018). *Selling race and god during GE13: a discourse-historical analysis of editorials and columns in mainstream Malay and English-language newspapers during the 13th general election in Malaysia* [Doctoral dissertation, Loughborough University]. EThOS. https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.763463

Nadzri, M. M. (2018). The 14th General Election, the fall of Barisan Nasional, and political development in Malaysia, 1957-2018. *Journal of Current Southeast Asian Affairs*, *37*(3), 139-171.

Nardone, C. (2018). 'Women and Work': a Cross-Linguistic Corpus-Assisted Discourse Study in German and in Italian. *Critical Approaches to Discourse Analysis across Disciplines, 10*(1), 167–186.

Partington, A. (2010). Modern Diachronic Corpus-Assisted Discourse Studies (MD-CADS) on UK newspapers: An overview of the project. *Corpora*, *5*(2), 83-108. DOI: https://doi.org/10.3366/cor.2010.0101

Partington, A. S. (2013). A corpus analysis of political language. In C. Chapelle (Ed.), *The Encyclopaedia of applied linguistics* (pp. 1–8). Wiley-Blackwell Publishing.

Partington, A. S. (2018). Welcome to the first issue of the Journal of Corpora and Discourse Studies. *Journal of Corpora and Discourse Studies*, *1*(0), 1–7. DOI: https://doi.org/10.18573/jcads.19

Prihantoro. (2022). A Collocation Analysis of 'Energy' in Brown Family Corpus. *E3S Web Conf.*, *pp. 359*, 1–6. https://www.e3sconferences.org/articles/e3sconf/abs/2022/26/e3sconf_icenis2022_03012/e3sconf_icenis2022_03012.html

Puyok, A. (2021). The fall of Warisan in Sabah's election: telltale signs causes and salient issues. ISEAS-Yusof Ishak Institute. Retrieved 7 September 2022, from https://www.iseas.edu.sg/wpcontent/uploads/2021/01/ISEAS_Perspective_2021_8.pdf

Rayson, P. (2003). Matrix: A Statistical Method and Software Tool for Linguistic Analysis through Corpus Comparison. PhD thesis, Lancaster University.

Reuters Institute Digital News. (2020). *Country and Market Data Asia Pacific* (Report No. 3.34). Reuters Institute Digital News Report 2020. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf

Rogers, T. (2019). "Differences Between Broadsheet and Tabloid Newspapers." ThoughtCo, 15 July 2019, thoughtco.com/broadsheet-and-tabloid-newspapers-2074248

Welsh, B. (2021). Chapter 11 A Holistic Society-Centred Analysis of the Sabah Election 2020 Results: Voters, Voting and Trajectories of Survival and Change. In Sabah from the Ground: The 2020 Elections and the Politics of Survival (pp. 231–264). Singapore: ISEAS Publishing. https://doi.org/10.1355/9789814951692-015

Wong, C. H. (2020). Sabah State Election 2020: Did the Electorate Vote for Change? Platform. *A Journal of Management and Humanities, 3*(2), 2–15.

APPENDIX A

SAMPLE KEYWORDS LIST TAKEN FROM #LancsBox6.0

Type	Frequency 1 (The Star Sabah Election News)	Dispersion 1 (The Star Sabah Election News)	Frequency 2 (Malaysiakini Sabah Election News)	Dispersion 2 (Malaysiakini Sabah Election News)	▼ Statistic
dr	782.000000	0.659967	346.000000	1.514728	832.5930218334775
tan	278.000000	0.695834	2.000000	18.428257	687.5113864140174
sri	276.000000	0.663209	46.000000	7.010527	470.8152765385241
today	26.000000	3.489865	960.000000	0.895204	453.8404217915599
noor	701.000000	0.729466	609.000000	1.617026	378.89293338161076
hisham	687.000000	0.729827	591.000000	1.648413	376.481658319232
published	3.000000	12.187470	609.000000	0.498053	368.02745873961624
total	438.000000	1.191406	296.000000	2.003741	322.2400240827788
cases	1733.000000	0.893158	2574.000000	1.478948	303.5241537709752
pm	7.000000	7.791108	540.000000	0.707423	296.08394923857236
datuk	115.000000	2.043205	3.000000	14.294892	267.6930587723849
cluster	849.000000	1.271984	1087.000000	2.030930	224.4924056322136
tuesday	121.000000	2.436775	28.000000	5.328973	183.32911384824132
friday	114.000000	2.395812	35.000000	6.639796	151.5730777846029
says	151.000000	1.415752	73.000000	6.004444	150.57048434157733
new	786.000000	1.069284	1142.000000	1.912941	147.44188814745985
thursday	90.000000	2.658227	18.000000	7.023478	144.29178326744693
country's	102.000000	1.918085	28.000000	5.046291	143.33708677059747
percent	0	0	212.000000	3.041261	138.63894894992083
individuals	221.000000	2.055526	172.000000	3.190247	138.18159010639073
were	793.000000	0.989403	1194.000000	1.217207	132.90471018495333
recorded	264.000000	1.457576	242.000000	2.250426	131.95213094077067
negri	57.000000	2.515758	3.000000	14.412695	123.70653831379906
wednesday	85.000000	2.886609	23.000000	6.108306	120.256905421504
number	359.000000	1.292920	414.000000	1.673153	119.86278136773765
said	1418.000000	0.474502	2534.000000	0.574294	119.25149656499559
monday	104.000000	2.349497	42.000000	4.633931	117.84561140206887
students	5.000000	10.654936	223.000000	4.117636	110.51283525750038
sept	381.000000	1.446514	478.000000	1.984693	105.73848063791905
sunday	85.000000	3.114636	30.000000	5.779169	104.67972126358633

APPENDIX B

SPILL-OVER FROM SABAH STATE ELECTION TO OTHER STATES IN MALAYSIA KINI (CONTINUED)

21	OCT_04_4_ COVID19 SABAH MK.txt	screening: One case Labuan (Five cases) screening	**of**	arrivals from high-risk areas in Sabah: Five
22	OCT_04_ SABAH AND KEDAH_MK.txt	from Sabah. This means that the number	**of**	cases with a travel history to Sabah since
23	OCT_04_4_293 NEW COVID19_MK.txt	Sabah: Two cases Perak (Two cases) screening	**of**	arrivals from high-risk areas in Sabah: Penang
24	OCT_04_4_SABAH STATES_MK.txt	One case Negeri Sembilan (Two cases) screening	**of**	arrivals from high-risk areas in Sabah: Two
25	OCT_05_14_OUTLET STAFF_MK.txt	the surge in cases of Covid-19, the majority	**of**	who were traced back to Sabah, which
26	OCT_06_2_ SABAH POLLS_MK.txt	235 new cases reported have a history	**of**	travel to Sabah. Although the sharp hike
27	OCT_07_5_ LINKED TO NEW_MK.txt	cases today: - Two cases during the screening	**of**	travellers returning from Sabah. - One case during
28	OCT_08_ EXPOSURE TO COVID19_MK.txt	of yesterday, there have been 283 cases	**of**	Covid-19 has been linked to travel to Sabah since
29	OCT_08_6_ BY ELECTION_MK.txt	election. In addition, 301 Covid-19 cases outside	**of**	Sabah, to date, have been linked to
30	OCT_08_10_COVID19 DEATHS_MK.txt	screening (two cases). Johor reported five cases,	**of**	which three involved travellers returning from Sabah
31	OCT_08_10_ SCREENING_MK.txt	Sarawak cases were detected through screening	**of**	travellers returning from Sabah (one case) and
32	OCT_08_10_ NEW CASES_MK.txt	in Perak today were detected through screening	**of**	travellers from Sabah (two cases) and other
33	OCT_10_2_ SABAH POLLS_MK.txt	On 8 October, there were 283 cases	**of**	Covid-19 has been linked to travel to Sabah since
34	OCT_11_2_ SECOND HIGHEST_MK.txt	date, the number of Covid-19 cases outside	**of**	Sabah, but traced to Sabah, stands at
35	OCT_12_1_ DIE COVID19_MK.txt	On 8 October, there were 283 cases	**of**	Covid-19 has been linked to travel to Sabah since
36	OCT_13_3_UNAVOID ABLE_MK.txt	the country. There have been 394 cases	**of**	Covid-19 has been linked to travel to Sabah since
37	OCT_16_2_COVID19 FLOODS_MK.txt	back in 1983. On the increasing cases	**of**	Covid-19 linked to the Sabah state election,
38	OCT_16_4_ BREACH 6000 MARK_MK.txt	recent travel from Sabah. The total number	**of**	cases that involved recent travel to Sabah
39	OCT_17_2_SABAH ICUS_MK.txt	said. Malaysia recorded 869 new cases today,	**of**	which 451 came from Sabah.
40	OCT_21_11_FORTNIG HT DEATH_MK.txt	4 Labuan (9 new cases) Other types	**of**	screening- 2 Returnees from Sabah- 2 Bina
41	OCT_22_16_ SPIKE IN LABUAN_MK.txt	Close contact screening- 55 cases - Other types	**of**	screening- 10 cases - Returnee from Sabah- 8