

Review Article

A Systematic Literature Review on Corpus-based Studies in Malaysian Media Discourse

**Nur Farhana Nasri^{1,2*}, Nor Fariza Mohd Nor¹,
and Anis Nadiyah Che Abdul Rahman¹**

¹*Centre for Research in Language and Linguistics, Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia*

²*Academy of Language Studies, Universiti Teknologi MARA, Perak Branch, Tapah Campus, 35400 Tapah Road, Perak, Malaysia*

ABSTRACT

Media discourse plays a vital role in shaping public perceptions, constructing ideologies, and influencing societal narratives. However, the study of media discourse in the Malaysian context has been relatively limited, particularly through a corpus-based lens. This review systematically identifies the research trends in Malaysian media discourse using a corpus-based approach. The included papers were extracted from two databases, namely Scopus and Web of Science (WoS), with 23 articles selected from 122 papers published between 2019 and 2024 based on specific exclusion and inclusion criteria. First, the significant findings show that there was underutilisation of advanced corpus tools and general corpora. Second, there was limited theoretical engagement within CDA frameworks. Third, the thematic analysis revealed a lack of focus on socioeconomic themes despite their societal relevance in Malaysia. These findings highlight ongoing gaps in Malaysian media discourse and corpus-based studies. Generally, this review provides insights for linguists

and researchers to understand the diverse portrayal of Malaysian media from multiple perspectives while identifying key gaps and directions for future research. To address these gaps, future studies are recommended to explore underrepresented socioeconomic issues and incorporate more diverse corpus designs and advanced analytical tools in Malaysian media discourse.

ARTICLE INFO

Article history:

Received: 04 September 2025

Accepted: 30 May 2026

Published: 30 June 2026

DOI: <https://doi.org/10.47836/pjssh.34.3.12>

E-mail addresses:

nfarhananasri@gmail.com (Nur Farhana Nasri)

fariza@ukm.edu.my (Nor Fariza Mohd Nor)

aniscar@ukm.edu.my (Anis Nadiyah Che Abdul Rahman)

* Corresponding author

Keywords: Corpus approach, Malaysia, media discourse, systematic review, thematic analysis

INTRODUCTION

Mass media refers to a medium that carries mass communication to a large audience. Mass media constitutes “any communication channel used to simultaneously reach a large number of people, including radio, TV, newspapers, magazines, billboards, films, recordings, books, and the Internet... [as well as] the new category smart mass media which include smartphones, smart TVs, and tablets” (Wimmer & Dominick, 2012, p.2). The way issues are framed in news articles, TV news, and political talk shows can affect how people understand them. Media messages would be picked by the audience, which will eventually influence their opinions (Entman, 1993).

The connection between language and media is undeniable, as language is used to convey messages within the media. Discourse-level analysis plays a crucial role in identifying distinctive features and linguistic practices within news language. Examining the media environment enables researchers to analyse data and gain a better understanding of the patterns, which deepens people’s knowledge of both media practices and discourse strategies (Cotter, 2001). Media discourse has evolved beyond the analysis of media texts. It also includes broader social practices that involve interdisciplinary linguistic perspectives. Therefore, linguists are increasingly interested in contributing to media studies by applying discourse-analytical methods to understand how language meaning is shaped by context (Hua, 2022).

One of the main concerns in critical linguistics is media discourse (Fairclough, 1995). Fairclough (1995) argues that the media can be influenced by powerful societal institutions. As a result, the media constructs “representations of reality” instead of “reality”. He emphasises that the importance of media analysis in analysing how media content is “read” by different audiences is just as vital as analysing the content itself. Various linguistic approaches, including critical discourse analysis (CDA), corpus-assisted discourse studies (CADS), appraisal analysis, and stance analysis, have been widely explored to examine how media texts can unveil underlying ideological meanings and linguistic patterns. Combining corpus and discourse approaches has also been proven to be an effective method for analysing news texts (Baker & McEnery, 2005; Xin & Gao, 2019; Zhao & Zhang, 2019).

With the emergence of digital media, newspapers have become the primary medium for disseminating information. In the Malaysian context, media operations have long been shaped by structural constraints such as censorship laws and political ownership. As Kwan and Ying (2023) point out, the government’s extensive influence has led to limitations on how information is framed, which can distort public opinion. In the same vein, Gomez et al. (2018) highlight how ties between traditional media outlets and political actors raise questions about the credibility of news coverage. These structural realities emphasise the importance of analysing

Malaysian media discourse, especially through corpus-based approaches, as they allow researchers to systematically uncover patterns of representation, ideological framing, or media bias that may arise in large bodies of text.

Many studies have examined the diverse portrayal of Malaysian media discourse, reflecting a strong interest among scholars in understanding how language shapes public narratives. In recent years, corpus-based approaches have become increasingly significant, offering empirical and replicable methodology for analysing large amounts of media texts. Despite the increase in corpus-based research within Malaysian media discourse, there is still a lack of systematic reviews that compile findings, methodologies, and research trends in this area. To the researcher's knowledge, no prior systematic literature review has specifically examined corpus-based studies in Malaysian media discourse. Therefore, this paper aims to review 23 studies that employ corpus-based approaches focusing on Malaysian media discourse for the last six years (2019-2024). The specific research questions are: 1) What are the research trends in Malaysian media discourse studies that employ a corpus-based approach in the last six years (2019 - 2024)? 2) What are the potential gaps and under-researched areas in the existing body of literature? The findings of this review may benefit corpus linguists, discourse analysts, and Malaysian media scholars by offering more knowledge on research trends, methodological developments, and

areas that need further investigation in corpus-based studies, specifically within the Malaysian media context.

METHODOLOGY

PRISMA, or Preferred Reporting Items for Systematic Reviews and Meta-Analyses, was first introduced in 2009, aiming to evaluate and synthesise studies systematically and transparently through a standard reporting guideline. The PRISMA 2020 guidelines, which represent a modification from the original 2009 version, comprise a 27-item checklist intended for systematic reviews assessing the impact of health interventions, regardless of the study designs involved (Page et al., 2021). However, the designed checklist items are also applicable to systematic reviews examining other domains, including social and educational fields. Notably, PRISMA has also been adopted in language and linguistics research (Bailey et al., 2020; Farrús, 2023; Masullo et al., 2023). In general, the PRISMA framework (Moher et al., 2009) highlights several steps in selecting papers: 1) identifying sources of information, 2) selecting studies, 3) conducting data collection, 4) determining eligibility criteria, and 5) selecting relevant data items.

Determining the right research questions helps researchers in selecting articles, extracting, and reporting data (Xiao & Watson, 2019). The formulation of research questions for this study was based on PICO (Lockwood et al., 2015). PICO is a widely used mnemonic to help researchers develop

research questions for systematic review. PICO is based on three main elements: P (Population or Problem), I (Interest), and Co (Context). Thus, this paper aims to examine 23 past studies to learn about emerging research trends, preferred topics, and theoretical/methodological approaches in corpus-based studies (Interest) in Malaysian media discourse (Population), published from 2019 until 2024 (Context).

Due to the subject-specific nature of databases and their tendency to index only a portion of the literature, it is recommended to search more than one database when conducting SLR studies (Papaioannou, 2010). Therefore, in the present study, Scopus and Web of Science (WoS) were selected for the review process. These databases are both comprehensive and robust, covering a wide range of disciplines and fields. WoS, owned by Clarivate, is the world's oldest research database and served as the sole source of bibliographic data for 40 years until Scopus was launched by Elsevier in 2004 (Baas, 2020). WoS comprises several citation indexes that include 254 subject areas and cover 34000 journals from the arts and humanities, sciences, and social sciences (Birkle et al., 2020). Conversely, Scopus is a user-friendly, multidisciplinary database spanning life sciences, humanities, engineering, and business, indexing nearly 42000 journals. Scopus is recognised for publishing high-quality research publications, with journals needing to meet rigorous quality standards to maintain high academic standards for researchers worldwide.

The Systematic Review Process

Identification

The systematic review process for identifying relevant articles consisted of three stages. The first stage is the identification of keywords, followed by the process of searching for synonyms, related terms, and variations using dictionaries and online thesauruses, asking experts' opinions, and referring to past studies. Based on this process, search strings on Scopus and Web of Science (WoS) databases were developed in April 2025 (Table 1) after all the keywords were determined. These keywords were processed using search functions such as field code functions, phrase searching, wildcards, truncation, and Boolean operators. Based on the searching process, a total of 122 potential articles were identified from both databases.

Screening

The first stage of the screening process was to remove duplicate articles. While 27 duplicated articles were eliminated, a total of 52 articles were excluded based on several inclusion and exclusion criteria determined by the researchers in the second stage. The first criterion was the type of literature, focusing on journals (research articles) containing empirical data. Therefore, other types of articles, such as systematic review articles, review articles, meta-analyses articles, meta-synthesis articles, book series, book chapters, and conference proceedings, were excluded.

It should be noted that the review only includes articles published in English, which may have contributed to a relatively small pool of selected data. Although accepting only English articles helps ensure consistency in the selection and analysis process, it may limit the findings' representativeness within Malaysia's multilingual media landscape. Therefore, relevant studies published in Malay or other languages that could offer different thematic foci, media representations, or methodological perspectives are not fully represented in this review. Future reviews

could benefit from including multilingual studies to gain a broader understanding of corpus-based research in Malaysian media. Additionally, the chosen timeline for the review was between 2019 and 2024, considering that the number of published papers during this period was sufficient to provide a representative review. Thus, articles published before 2019 were not included (Table 2). This process resulted in 43 articles remaining for evaluation in the next stage.

Table 1
Search string used in the selected database

Database	String
Scopus	TITLE-ABS-KEY (("media discourse*" OR "news discourse*" OR "news coverage*" OR "media report*" OR "news report*" OR "media representation*" OR "media text*" OR "media narrative*" OR "journalistic discourse*" OR "press discourse*" OR "media portrayal*" OR "media framing*" OR "news content*" OR "media language*") AND ("corpus-based" OR "corpus linguistics*" OR "corpus-driven" OR "corpus-assisted" OR "discourse analysis" OR "critical discourse analysis" OR "linguistic analysis" OR corpus*) AND (malaysia OR malaysian))
WoS	TS= (("media discourse*" OR "news discourse*" OR "news coverage*" OR "media report*" OR "news report*" OR "media representation*" OR "media text*" OR "media narrative*" OR "journalistic discourse*" OR "press discourse*" OR "media portrayal*" OR "media framing*" OR "news content*" OR "media language*") AND ("corpus-based" OR "corpus linguistics*" OR "corpus-driven" OR "corpus-assisted" OR "discourse analysis" OR "critical discourse analysis" OR "linguistic analysis" OR corpus*) AND (malaysia OR malaysian))

Table 2
The inclusion and exclusion criteria

Criterion	Inclusion	Exclusion
Literature Type	Journal (articles with empirical data)	Review articles, book series, book, chapter in book, conference proceedings, etc.
Language	English	Non-English
Timeline	2019-2024	2018 and earlier

Eligibility

A total of 43 articles were prepared for analysis during the eligibility process. During this stage, the titles, abstracts, and main contents of all articles were carefully examined to ensure that they met the inclusion criteria and were suitable for the review. The articles were manually screened to confirm the actual use of corpus-based approaches and corpus analytical tools. Consequently, 20 articles were excluded because they had inaccessible full texts, were not based on empirical data, or did not focus on Malaysian media outlets.

Furthermore, studies that employed purely qualitative discourse analysis without the use of corpus compilation, corpus software, or quantitative corpus techniques were also excluded from the review. Finally, a total of 23 remaining articles were ready for analysis. The flowchart of the study selection process is presented based on Moher et al.'s (2009) reporting items for systematic reviews and meta-analyses (PRISMA) in Figure 1.

This study employed a two-step analysis as suggested by Ahmed and Matthes (2017).

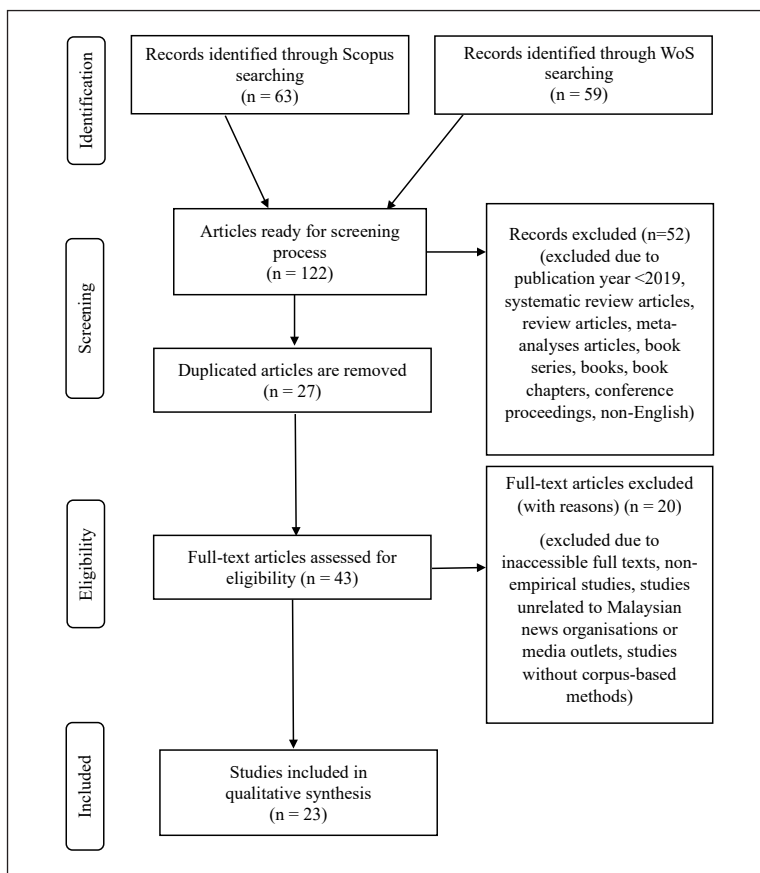


Figure 1. Flowchart of the study selection process (adapted from Moher et al., 2009)

The first step was a detailed quantitative analysis to code and categorise key characteristics of the selected studies, understand the research trends, and make comparisons. The studies were coded into the subsequent subtopics: 1) year of publication, 2) corpus involved, 3) corpus size, 4) instrument/tool used, and 5) theoretical/methodological approaches. The second step was the qualitative analysis to identify significant themes from the reviewed literature. Each study was grouped based on its primary thematic focus and organised under the same thematic categories.

RESULTS

The main objectives of this systematic review are i) to find out the research trends in Malaysian media discourse studies that employ the corpus-based approach in the last six years (2019 - 2024) and ii) to identify the potential gaps and under-researched areas in the existing body of literature.

To achieve the first objective, the analysis of the research trend from 2019 until 2024 was presented in Table 3. It summarises the author and publication year, type of corpus, instrument/tool used, and the theoretical/methodological perspective employed in each study selected.

Published Studies Trend

Figure 2 shows the number of published studies using the corpus approach from 2019 to 2024. The average number of publications over the six years remains relatively low, with an overall mean of 3.83. The figure shows a fluctuating trend of research

studies as the number of studies, which was consistent with 4 studies in 2019 and 2020, began to decline, with just 2 studies in 2021. Nevertheless, the number returned to 4 again both in 2022 and 2023. Fortunately, in 2024, the number of published studies increased slightly to 5 papers. The pattern suggests that though corpus-based study is gaining more attention among researchers, this field of study has been scarcely examined. Consequently, it highlights a significant opportunity for this field to be further explored in the future.

Corpus Used

Figure 3 illustrates the distribution of corpus types used in the reviewed studies. Most studies (83%) used a target corpus, typically compiling news reports sourced from Malaysian newspapers for their analysis. On the contrary, only 17% of the studies (Abas & Kaur, 2019; Abdul Malik et al., 2022; 2023; Wang & Liang, 2024) adopted a mixed corpus approach, incorporating a target corpus and a reference corpus. Notably, in the studies where a reference corpus was used, the British National Corpus (BNC) served as the comparative baseline, thus suggesting that these authors were interested in comparing British English and the English language used in the Malaysian newspapers.

In addition, Figure 4 shows the percentage of the corpus used based on fields in the selected literature. It is clearly stated that most authors employed news reports from the health field (35%).

Table 3
Descriptive overview of reviewed studies

No.	Author	Corpus	Corpus size	Theories/ Methods	Instruments/ Tools
1	Chang and Abdul Rahim (2021)	News reports on Covid-19 during MCO 1.0 (1 st - 30 th April 2020)	1299 articles (971283 words)	Corpus Linguistics, Semantic prosody (Louw, 2000)	AntConc version 3.5.9
2	Asad et al. (2019)	News reports on social actors' performances (3 rd May - 9 th May 2018)	21 articles (8395 words)	Fairclough's (1995) Critical Discourse Analysis, Halliday's Systematic Functional Linguistics (SFL)	Excel spreadsheet
3	Wang and Liang (2024)	British National Corpus (BNC), news reports from Malay-, English-, and Chinese-language newspapers (16 th October 2022 - 16 th October 2023)	192 articles (138647 reports)	Corpus-Assisted Discourse Analysis (CADS)	AntConc 4.2.4
4	Ong and Abdul Rahim (2019)	News articles on GST from GST Malaysia Info Website	10 articles (3172 words)	Corpus Linguistics, Fairclough's (1995) Critical Discourse Analysis	AntConc 3.4.4
5	Ismail et al. (2020a)	Orang Asli corpus (February 2003 - February 2019)	68000 words	Corpus-Assisted Critical Discourse Analysis (CACDA)	WordSmith Tools 6.0
6	Wu et al. (2022)	Hard news and opinion pieces on world's economy, politics, culture etc. about China (January 2012 - December 2021)	137475 articles (65 million words)	Sentiment Analysis, Corpus Linguistics, Discourse Analysis	WordSmith Tools 8.0
7	Yoong and Lee (2023)	News reports (dated May - October 2018)	84 articles	Critical Discourse Analysis and Jeffries's (2010) Critical Stylistics model	Search queries

Table 3 (continued)

No.	Author	Corpus	Corpus size	Theories/ Methods	Instruments/ Tools
8	Mohd Nor et al. (2021)	Corpus of Malaysian Online News Articles of Mental Health (1 st January 2020 - 31 st December 2020)	585 articles (320509 words)	Corpus Linguistics, Discourse Analysis	AntConc 3.5.8
9	Mohd Nor and Zulcafli (2020)	News reports on Covid-19 (1st - 31st March 2020)	1018 articles	Corpus Linguistics, Discourse Analysis	AntConc 3.5.8
10	Ismail et al. (2020b)	Malaysian Sports News (MSN) Corpus (1st - 30th November 2013 and 1st - 31st March 2014)	7640 articles (approx. 3000000 words)	Corpus-Assisted Discourse Analysis (CADS)	WordSmith Tools 6.0
11	Roslani and Jamal (2024)	News articles of Nur Sajat (23rd February 2021 - 23rd February 2022)	29 articles	van Leeuwen's Representation of Social Actors (2008) of Critical Discourse Analysis	Search queries
12	Abas et al. (2023)	Malaysian Corpus of Vaccination (M-CorVax) (March 2020 - June 2021)	1514 articles	Corpus-Assisted Discourse Analysis (CADS) with Discourse-Historical Approach (DHA)	#LancsBox v.6.x.
13	Murugaiah et al. (2024)	News reports on people with mental illness (PwMI)	24 articles	van Dijk's (1980) Theory of Semantic Macrostructures, van Dijk's (1998) Ideological Square model of Critical Discourse Analysis	Search queries
14	Abdul Malik et al. (2022)	British National Corpus (BNC) 2014, COVID-19 Vaccine Related News Reports in Malaysian English Newspapers Corpus (CocVac-MEN Corpus)	538 articles (178675 words)	Corpus Linguistics	#LancsBox 5.0.

Table 3 (continued)

No.	Author	Corpus	Corpus size	Theories/ Methods	Instruments/ Tools
15	Abdul Malik et al. (2023)	British National Corpus (BNC) 2014, COVID-19 Vaccine Related News Reports in Malaysian English Newspapers Corpus (CocVac-MEN Corpus)	538 articles (178675 words)	Corpus Linguistics	#LancsBox 5.0
16	Joharry et al. (2023)	News reports 30 days before and 30 days after Sabah election day on 26 September 2020	996 articles (390868 words)	Corpus-Assisted Discourse Analysis (CADS)	#LancsBox6.0
17	Abas and Kaur (2019)	British National Corpus (BNC), news reports from sports section	1311 articles (485862 words)	Corpus Linguistics, Discourse Analysis	Not mentioned
18	Mohd Ali et al. (2024)	Newspaper reports on mental illness among Malaysians (January 2017 - December 2021)	43 articles	Reisigl and Wodak's (2009) discursive strategies of Critical Discourse Analysis	Search queries
19	Khalid et al. (2020)	News reports concerned with the news program of direct hiring (SMO)	29 articles	Discourse-Historical Approach (DHA) framework of Critical Discourse Analysis (Reisigl & Wodak 2001)	NVivo 12 Pro
20	Selvaraj and Sandaran (2024)	News reports on aging or aging population (May 2022 - May 2023)	83 articles	Discourse Analysis	Search queries
21	Asad (2022)	News reports on social actors i.e. Dr Mahathir, Najib Razak, Pakatan Harapan (PH), and Barisan Nasional (BN) (17 th - 18 th August 2018)	8 articles (5721 words)	Fairclough's (1995) model of CDA, Halliday's (1985) Systemic Functional Linguistics (SFL)	Excel spreadsheet

Table 3 (continued)

No.	Author	Corpus	Corpus size	Theories/Methods	Instruments/Tools
22	Taufek et al. (2022)	News reports on climate change (2015 - 2021)	922 articles (804289 words)	Corpus Linguistics, Discourse Analysis	AntConc 4.0.5
23	Hajimia et al. (2019)	News reports on social and environmental issues (March - October 2016)	30 articles (10500 words)	Corpus-assisted Corpus Analysis (CACA)	AntConc

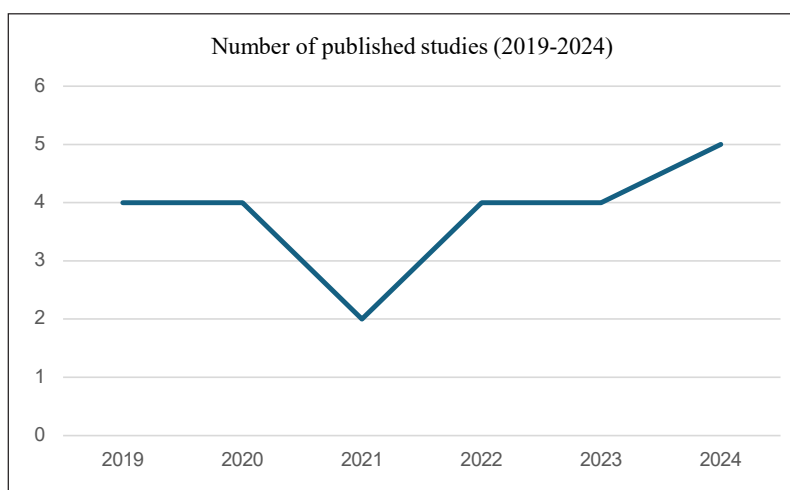


Figure 2. Trend of corpus-based studies in Malaysian media discourse (2019 - 2024)

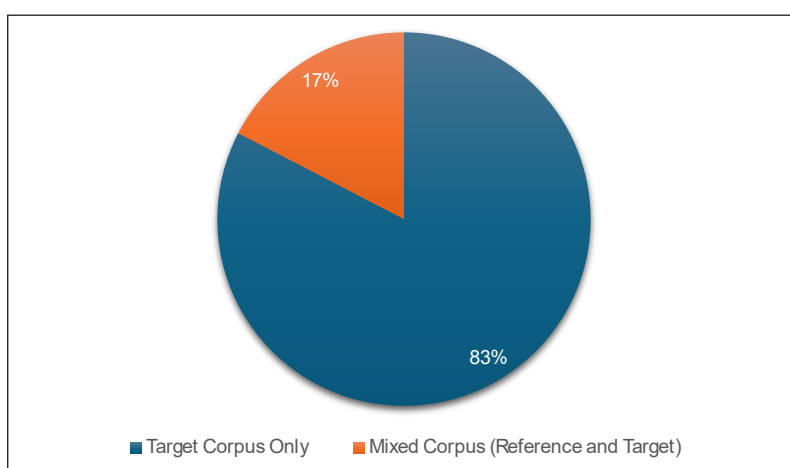


Figure 3. The percentages of corpus types in the selected literature

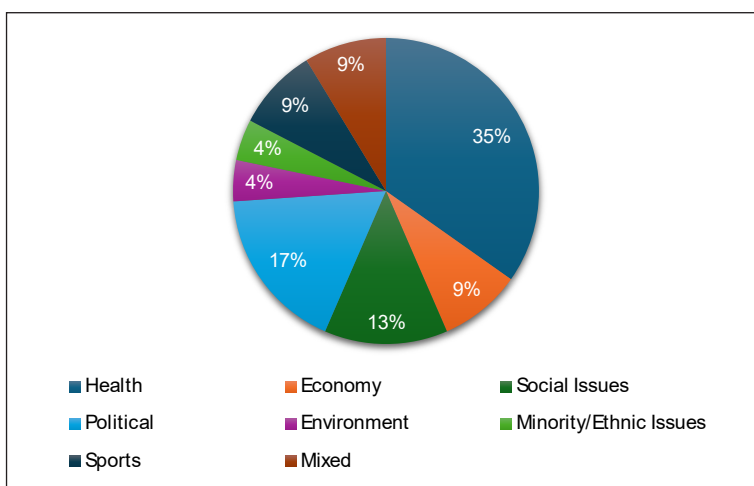


Figure 4. The percentages of the corpus used based on the fields in the selected literature

For example, Abdul Malik et al. (2022) and Abas et al. (2023) built their corpus from newspaper reports called CocVac-MEN Corpus and Malaysian Corpus of Vaccination (m-CorVax), respectively. Likewise, Mohd Nor et al. (2021) and Murugaiah et al. (2024) examined news articles concerning mental health issues, thus contributing to health as the most frequently studied topic in corpus-based media discourse. In the meantime, the second-highest corpus field is from political discourse, such as Asad (2022), who examined political social actors, and Joharry et al. (2023), who investigated the 2020 Sabah state election. Social issues covered 13% of all studies, where topics such as transgender (Roslan & Jamal, 2024), ageing (Selvaraj & Sandaran, 2024), and migrant women (Yoong & Lee, 2023) were explored. In contrast, economy, minority/social issues, and mixed fields shared a similar percentage with 9%, whereas the lowest percentage, which is 4%, is shared by two other fields: environment and sports.

Instruments/Tools Used

As illustrated in Figure 5, the most frequently used instrument in corpus-based studies of Malaysian media discourse between 2019 and 2024 was AntConc with 30% (n=7), followed by search queries with 22% of the studies (n=5). Next, #LancsBox was used in 17% (n=4) while WordSmith Tools featured in 13% of the studies (n=3). Additionally, 8.7% (n=2) of the studies used a Microsoft Excel spreadsheet for manual data processing. A single study (4.3%) employed NVivo 12Pro, whereas another study (4.3%) did not specify the specific instrument or tools used.

To illustrate, studies that employed AntConc (e.g., Chang & Abdul Rahim, 2021; Wang & Liang, 2024) and #LancsBox tool (Abdul Malik et al., 2022; Joharry et al., 2023) mostly involved small to medium corpus data, while studies that used WordSmith Tools (e.g., Ismail et al., 2020b; Wu et al., 2022) analysed larger datasets.

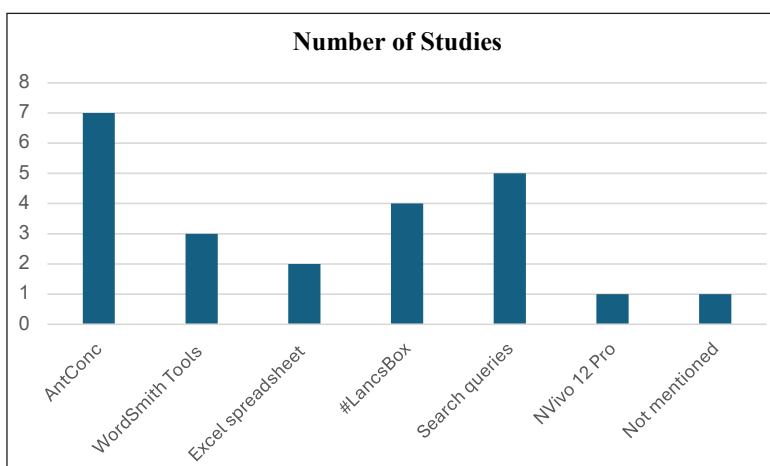


Figure 5. The instruments or tools used by authors in the selected literature

Meanwhile, Asad et al. (2019) and Asad (2022) utilised Excel spreadsheet software to manage and analyse news reports that had been downloaded manually, considering the small corpus sizes with only 21 in 2019 and 8 in 2022. Given the limited amount of data, the authors opted for manual processing and basic analysis for data compilation. On the other hand, Khalid et al. (2020) employed NVivo 12 Pro as a preliminary analysis for their qualitative study. They used the software's word frequency function to generate a word cloud. The use of this software enabled the researchers to obtain and illustrate broad findings within the data.

On the contrary, Selvaraj and Sandaran (2024), Mohd Ali et al. (2024), Murugaiah et al. (2024), and Roslani and Jamal (2024) relied on search queries to manually retrieve articles from online newspaper databases using keywords related to their studies. These studies were primarily qualitative and involved small datasets.

The authors prioritised keyword-based data collection suited for their interpretive frameworks. Interestingly, a study by Abas and Kaur (2019) did not specifically mention the analytical instrument they used, despite reporting concordance and collocation reporting as methods for determining the concept of theologizing in football news discourses.

Theoretical/Methodological Approaches

Figure 6 illustrates the theoretical and methodological approaches adopted in Malaysian media discourse studies from 2019 to 2024. The bar graph indicates that while some studies relied on specific theoretical frameworks, others focused on methodological tools to guide their data analysis. It is clearly shown that most studies (35%) employed Critical Discourse Analysis (CDA) (Khalid et al., 2020; Murugaiah et al., 2024; Mohd Ali et al., 2024; Roslani & Jamal, 2024).

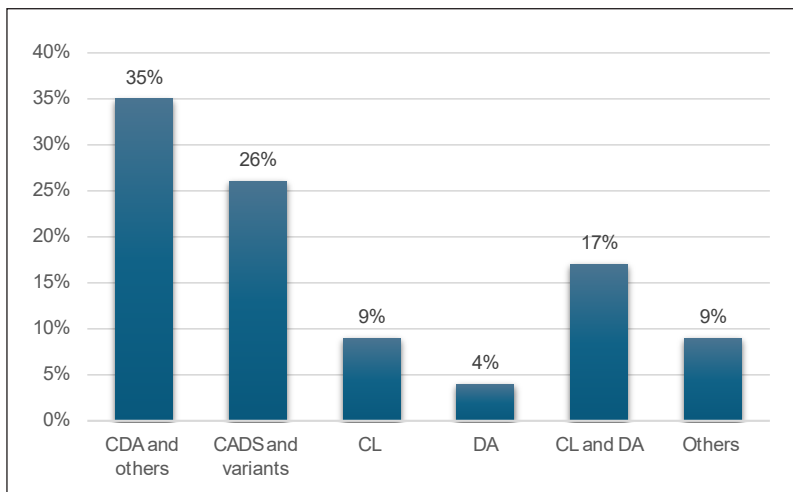


Figure 6. The theoretical and methodological approaches used in the reviewed studies

Some studies used CDA and paired it with other theories, such as Asad et al. (2019) and Asad (2022), who examined political social actors' performances in Malaysian elections using CDA and Halliday's Systematic Functional Linguistics (SFL). Another study conducted by Yoong and Lee (2023) combined CDA with Jeffries's (2010) Critical Stylistics model to unveil how a Malaysian newspaper represents Chinese migrant women, while Ong and Abdul Rahim (2019) used Corpus Linguistics (CL) as a method to facilitate preliminary analysis before the data was analysed using the CDA approach.

In addition, 26% of the studies employed Corpus-Assisted Discourse Analysis (CADS) and its variants. For instance, Wang and Liang (2024) employed the CADS approach using Fairclough's (1989) model to examine how Malaysian media portrayed Chinese modernity through language choices, syntactic structures, metaphors, and narrative strategies. Several

studies also utilised CADS, such as Ismail et al. (2020b), who discussed the emotional portrayal of male and female athletes in sports news discourse by analysing keywords, collocations, and evaluative language. Abas et al. (2023) employed corpus analysis and Wodak's Discourse-Historical Approach (DHA) to investigate media coverage of vaccination-related news. Similarly, Joharry et al. (2023) explored news reporting strategies used before, during, and after the Sabah state election by comparing significant keywords and collocates. Other variants of CADS were also evident in the review. For instance, Corpus-Assisted Critical Discourse Analysis (CACDA) was employed by Ismail et al. (2020a) to examine Malaysian media representation of Orang Asli, revealing how language patterns signified ideologies of power and powerlessness. Another variation was Corpus-assisted Corpus Analysis (CACA) employed by Hajimia et al. (2019) to examine the distributional patterns of

discourse markers in selected newspaper articles.

Rather than employing a specific theoretical framework, some studies adopted a data-driven corpus linguistic approach. For example, Abdul Malik et al. (2022, 2023) employed the CL method to examine how COVID-19 vaccination in news reports was framed, amounting to 9% of all studies. In addition, several studies (17%) combined CL with Discourse Analysis (DA) to further understand language patterns in selected news articles. Abas and Kaur (2019) analysed the linguistic connection between ‘faith’ and religion in football coverage, while Mohd Nor et al. (2021) and Mohd Nor and Zulcafi (2020) explored the media portrayal of mental health issues and Covid-19 in Malaysian newspapers. Similarly, a study by Taufek et al. (2022) also adopted CL and DA to conduct a thematic analysis of climate change discourse in Malaysian news reports.

Few studies (9%) fall under “Others”, with less commonly used methodological approaches. Some of these studies combined CL or DA with additional analytical approaches to offer more insights into media discourse. For example, Chang and Abdul Rahim (2021) integrated CL with semantic prosody to analyse media representation of Covid-19 in Malaysian media, while Wu et al. (2022) employed sentiment analysis with CL and DA to explore the representation of China in Malaysian news. In contrast, Selvaraj and Sandaran (2024) relied solely on DA to reveal dominant discourses related to ageing.

Thematic Analysis

The second research objective for this study is to explore potential gaps and under-researched areas in the existing body of literature. To fulfil the objective, this study carried out a thematic analysis of Malaysian media discourse, which employed a corpus-based approach. The data from the thematic analysis revealed several themes, which are illustrated in Table 4. Table 4 shows the identified themes from the analysed studies with the aim of achieving the second research objective.

Table 4 shows that a diverse range of topics were explored, reflecting the interdisciplinary nature of Malaysian media discourse research. Accordingly, nine themes were identified, with 8 out of 23 studies focused on media discourse of public policy and national campaigns. For example, Ong and Abdul Rahim (2019) examined news articles to investigate the implementation of goods and services tax (GST) in several mainstream newspapers. Using CL and Fairclough’s model of CDA, the study unveiled the discourse of anti-profiteering law enforcement, where mainstream newspapers significantly portrayed the success of GST through positive representation of the authorities and negative representation of perpetrators and offenders. The study revealed discursive strategies of *us vs them* in the reports, which shifted consumers’ anger towards unethical traders, thus shaping the public’s trust in the authorities’ credibility.

Table 4
 Overview of the qualitative thematic screening

No.	Themes	Author(s)/Year
1	Public Policy and National Campaigns	Chang & Abdul Rahim (2021) Ong & Abdul Rahim (2019) Mohd Nor & Zulcaffi (2020) Abas et al. (2023) Abdul Malik et al. (2022) Abdul Malik et al. (2023) Joharry et al. (2023) Khalid et al. (2020)
2	Political actors and ideological stance	Asad et al. (2019) Asad (2022)
3	Marginalised and vulnerable groups	Ismail et al. (2020a) Yoong & Lee (2023) Mohd Ali et al. (2024) Selvaraj & Sandaran (2024)
4	Linguistic features and stylistic patterns	Hajimia et al. (2019)
5	Environmental discourse and public engagement	Taufek et al. (2022)
6	Cultural and ideological framing	Abas & Kaur (2019)
7	International relations and cross-cultural interactions	Wang & Liang (2024) Wu et al. (2022)
8	Health and well-being in society	Mohd Nor et al. (2021) Murugaiah et al. (2024)
9	Gender and identity	Ismail et al. (2020b) Roslani & Jamal (2024)

Similarly, Mohd Nor and Zulcaffi (2020) studied verbs and noun collocates in a mainstream newspaper coverage of COVID-19. Their findings demonstrated how lexical items such as verbs (e.g., *spread*, *fight*) and nouns (e.g., *cases*, *outbreak*) collocate underscored the gravity of Covid-19. Furthermore, concordance analysis of the study also highlighted how the media discursively constructed the government's proactive stance in mitigating and curbing the pandemic, thereby encouraging public engagement

in the efforts. Khalid et al. (2020), on the other hand, explored media representation of the Online Maid System (SMO) as a direct hiring system in recruiting foreign domestic helpers. The study unveiled that mainstream and alternative newspapers showed contradictory representations. While the former depicted SMO as efficient and convenient, the latter doubted the system. The issue was largely framed from an economic and procedural perspective, but limited attention was given to the effects of the system on foreign domestic workers.

Both newspapers underscored the difference in ideological positions, particularly in the coverage of public policies and campaigns. Only 3 out of 8 studies incorporated both mainstream and alternative newspapers. This suggests that future research on media discourse could benefit from analysing both media types, as both may offer different interpretations in understanding public policy narratives in Malaysia.

While public policy and national campaigns emerged as the most researched themes, other areas received comparatively less attention. Only 4 studies focused on the representation of marginalised and vulnerable groups. While Yoong and Lee (2023) highlighted the media's role in shaping perceptions of foreign nationals, especially the migration of mainland Chinese women, the remaining 3 studies concentrated on vulnerable groups within the Malaysian context. Ismail et al. (2020a), for example, explored media representations of Orang Asli by analysing language patterns to uncover underlying ideologies of power and powerlessness between indigenous communities and authoritative figures. Another study by Selvaraj and Sandaran (2024) unveiled dominant discourses surrounding ageing population, aiming to contribute towards effective policies and initiatives for their well-being. On the contrary, Mohd Ali et al. (2024) examined the type of linguistic devices in representing mentally ill people, revealing a tendency towards negative portrayal that associates them with perpetrators. Despite these contributions, several under-researched areas remained, particularly from socio-

economic perspectives that are deeply intertwined within Malaysian society. It is crucial to bring to attention that issues such as income disparity, the marginalisation of communities, and the urban-rural divides have not been extensively explored in Malaysian corpus-based studies.

Several themes also received limited scholarly attention over the past six years, typically represented by one or two studies each. For instance, only two notable studies were conducted between 2019 and 2022 on political actors. Asad et al. (2019) examined the representation of social actors during Malaysia's general elections, while another study in 2022 focused specifically on the ideological portrayal of Dr. Mahathir Mohamad. Though these studies analysed news coverage from both mainstream and alternative newspapers, the review suggests that there is a limited focus on the political actor selection and media sources. Furthermore, studies on health and well-being remained limited in scope, as Mohd Nor et al. (2021) and Murugaiah et al. (2024) focused exclusively on mental health and mental illness. Similarly, research on international relations and cross-cultural interactions was primarily centred on China. For example, Wang and Liang (2024) investigated Malaysian media perspectives on Chinese modernisation, while Wu et al. (2022) explored sentiments towards China in Malaysian media. In the area of gender and identity, Ismail et al. (2020b) addressed the issue of masculinity, emotions, and gender through sports reporting, whereas Roslani and Jamal (2024) critically

analysed media representation of a local transgender woman.

Finally, the same lack of focus was observed in studies related to linguistic features and stylistic patterns, environmental discourse, and cultural and ideological framing, as there was only one study on each area based on the analysed thematic analysis. Although the reviewed studies explore a diverse range of topics in Malaysian media discourse, the limited number of studies within each indicates that many areas remain fragmented and underexplored.

DISCUSSION

Based on the distribution data of published studies, research in Malaysian media discourse using a corpus-based approach has not shown a consistent upward trend from 2019 to 2024. Instead, the number of studies fluctuated over the years. However, the slight increase in studies from 2023 to 2024 indicated a positive interest in this area. This review was exclusively based on newspaper reports. While newspapers play an integral role in framing the world around us (Copeland, 2003), they are often suppressed by social powers such as political institutions or interest groups that have the authority to control what events are covered in news reporting. News reports analysis enables scholars to unveil linguistic features, ideological framing, and the shaping of public opinion through selective reporting. Since this review focuses mainly on news reports, which are often archived, searchable, and relatively easy to compile and process, studies involving other media

and multimodal formats may be overlooked. It is crucial to highlight that other types of media, such as social media platforms and television broadcasts, also contribute to the public's perceptions. While these forms of media were not covered in the reviewed studies, previous work on other media, such as radio discourse (Chukwunalu et al., 2022; Valentine et al., 2018), demonstrates that the combination of spoken language and textual information impacts public interpretation in unique ways. Therefore, broader corpus-based studies that involve audiovisual elements, including multimodal, spoken, and digital media as data sources, could expand a deeper understanding of how media discourse evolves.

Additionally, Malaysian media discourse relies primarily on specialised corpora, which enables researchers to focus on specific areas within media, such as health news or social issues, and thereby allows for the comprehension of linguistic patterns in a particular domain or field. The use of specialised corpora remains inadequately explored (Sadjirin et al., 2018). Specialised corpora enable researchers to have greater control over topic selection and time frames and are highly context-specific while allowing studies to be tailored closely to Malaysian settings. Researchers are therefore encouraged to build their corpora (Baker, 2006) because a properly designed, specialised small corpus can offer relevant, significant findings regarding "specialised lexis and structures" (O'Keefe et al., 2007, p. 198) compared to a large corpus, which is not personalised

to fulfil researchers' requirements (Nesi, 2012). However, the current distribution reveals a clear corpus design gap in the limited use of general corpora. According to Leech (2002), a general corpus serves as a valuable reference, covering both written and spoken language over time. The use of large reference corpora in exploring media texts, particularly in CDA studies, reduces personal bias or intuition in how language is constructed. It serves as a linguistic baseline in determining typical language patterns when comparing specific texts (O'Halloran, 2010). Thus, researchers can differentiate between linguistic features that are distinctive in general Malaysian English and those that are ideologically marked or specific to media discourse, such as in news reporting.

Furthermore, AntConc and #LancsBox were the most frequent tools applied in most studies. Both offer similar core corpus functions such as keyword analysis, word frequency, collocations, and concordance analysis. They are accessible, user-friendly, and free, making them particularly attractive to researchers working with small datasets and constrained timelines. In comparison, WordSmith Tools was employed in studies with larger datasets, such as Ismail et al. (2020a), who analysed sixteen years' worth of news reports. Despite the usefulness of these tools, several methodological gaps emerge. A few studies relied on Excel spreadsheets and manual keyword searches. Though these tools are sufficient for narrowly focused and qualitative analyses, they may compromise the replicability, scalability,

and analytical rigour that are expected in corpus-based research. Noticeably, none of the reviewed studies used advanced corpus analysis tools such as Sketch Engine and CQPweb, which offer enhanced capabilities, large-scale, and robust corpus analysis (Gomide, 2020; Hardie, 2012; Kilgariff et al., 2014). Underutilisation of these tools may reflect methodological challenges within the Malaysian context, including a lack of technical training, restricted exposure to advanced corpus platforms, and limited accessibility. Sketch Engine, for instance, is often associated with sophisticated linguistic analysis and advanced users due to its extensive analytical features (Jabeen & Rashid, 2025). Therefore, limited engagement with advanced platforms as such may hinder the development of large-scale and methodologically diverse Malaysian media discourse studies.

Meanwhile, it can be inferred that CDA is the dominant theoretical approach across most of the reviewed studies. Most studies applied CDA broadly using Fairclough's 3D Model (Ong & Abdul Rahim, 2019; Wang & Liang, 2024). This preference may stem from foundational work by Chouliaraki and Fairclough (1999), who argue that media institutions that appear neutral and objective have the power to shape, mediate, and reinforce societal views and attitudes. In recent years, CADS has also emerged as a prominent and significant tool for linguistic research (Baker, 2015; Samaie & Malmir, 2017), reflecting a deeper understanding of how language functions in media by

combining quantitative and qualitative data. However, the review indicates a theoretical gap as CADS in Malaysian media discourse is scattered and insufficient, with a lack of engagement on specific tools within the CDA framework (e.g., social actors' analysis, transitivity). Only a handful of studies incorporated semantic analysis or semantic prosody despite their potential in uncovering evaluative meanings in news discourse.

Moreover, from the perspectives of corpus fields and thematic analysis, the findings reveal several thematic gaps in the reviewed studies. The review covered a range of subjects, including health, politics, and economics, but sports and environmental discourse remain underexplored. The thematic analysis also demonstrates that though the studies cover a variety of issues, the distribution of research focus remains uneven. Most studies concentrated on public policy and national campaigns, whereas fewer studies emphasised critical societal issues such as matters of migration, ageing population, and the representation of Orang Asli communities. However, socioeconomic issues that are closely intertwined within Malaysia (e.g., Ariffin et al., 2022; Abdul Latip & Othman, 2021), such as poverty, income inequality, and urbanisation, are understudied from a linguistic perspective. As manifested in other broader studies in media and communication studies, "issues to do with economic, class or related social inequalities (other than specific conceptualisation of gender, race,

and sexual orientation) do not feature prominently" (Preston, 2016, p.38). Given the media's role in facilitating public discourse, marginalised communities affected by economic challenges remain underexplored, highlighting a critical gap for corpus studies in the future.

Recommendations

This review highlights several gaps in the existing studies. First, although the focus of this review is on newspaper reports, it is crucial to note that other media forms, such as social media, TV broadcasts, and multimodal content, contribute significantly to media discourse. Several scholars (Cushion & Richard, 2013; Tiffen et al., 2014) argue that different forms of media vary in how they produce and disseminate news content. For example, social media tends to be unedited and less standardised compared to other language forms, such as news and novels (Brandtler, 2020), while multimodal discourse requires an interdisciplinary perspective where language is a vital mode that works alongside images to create meaning (Kress, 2012).

Secondly, utilising a general corpus in corpus study as a reference corpus helps identify linguistic features in news discourse. General corpora, such as the British National Corpus (BNC), which is designed to represent a wide range of modern British English, are very large and thus appear as a useful resource and a 'benchmark' in determining typical language when compared with a specialised corpus (McEnery, 2006).

Third, future studies could adopt advanced corpus tools such as Sketch Engine and CQPweb, which offer a wide range of sophisticated analytical functions for in-depth exploration of language patterns in media studies. The selection of appropriate tools should align with the study's aims, the features required, and the nature of the data being analysed. At the same time, these tools should also maintain a level of usability and flexibility, thus allowing researchers to manage, query, and visualise their data in ways that suit their specific analytical needs (Gomide, 2020). Since most studies rely on tools such as AntConc, WordSmith Tools, and #LancsBox, future studies should focus on optimising search strategies and making full use of the tools' functionalities. This includes using precise and linguistically informed search queries and exploiting tools' functions such as collocations and dispersion plots for a deeper understanding of discursive patterns.

Fourth, future research should adopt a more integrated CDA approach within CADS frameworks such as socio-cognitive approaches (van Dijk, 1991), social actor representation (van Leeuwen, 2008), and discourse-historical approach (Reisigl & Wodak, 2016) that could also significantly enrich the interpretive depth and critical interpretation of media texts. This review also promotes CADS as a methodological approach to uncover evaluative meanings in media discourse. Due to criticisms of CDA, where data collection is subjected to the researcher's interpretation and its demanding and time-consuming process, Baker (2006) suggested the integration of a

corpus-assisted approach. CADS minimises such flaws, as adding corpus linguistics helps complement the qualitative focus of CDA (Fairclough, 2013). It offers a bottom-up approach to process large data and offer reliable findings based on statistics (Baker, 2006). Nevertheless, qualitative analysis is still necessary to understand the context of the data. Therefore, combining both statistical data from corpus linguistics with the interpretative analysis of CDA helps to balance out the strengths and limitations of each approach.

Finally, given that the media landscape in Malaysia is economically and structurally influenced by political parties, it is pertinent to examine media representation from both mainstream and alternative newspapers. Studying both types of media can provide a more inclusive and comprehensive understanding of media discourse in Malaysia. Although the analysis of thematic trends explores several themes, topics involving marginalised communities and socio-economic inequalities require further exploration. Malaysia is a multiracial and socio-economically diversified country, where newspapers serve not only as sources of information but also as instruments for promoting political ideologies (Mat Yakim et al. 2019). Such issues are deeply embedded in Malaysia's national development plans and policies, yet they remain underrepresented in linguistic and media analyses. Therefore, future research should prioritise these themes to uncover how the media shapes public perceptions and influences policy responses concerning socio-economic challenges in Malaysia.

CONCLUSION

The main purpose of this study is to systematically review past research on Malaysian media discourse using a corpus-based approach from recent years (2019-2024). This review highlights several gaps within corpus-based research in Malaysian media discourse. Most studies relied on specialised corpora built from Malaysian newspaper reports, in alignment with the studies' research aims and objectives. Tools such as AntConc, search queries, and #LancsBox were commonly used and selected based on the nature of the study and data availability. In terms of methodological framework, Corpus-Assisted Discourse Analysis (CADS) and Critical Discourse Analysis (CDA) remain dominant in uncovering discursive patterns and ideologies in the media. Furthermore, over a third of the studies were mainly on public policy and national campaigns. In contrast, fewer studies examined critical societal concerns, such as the representation of marginalised groups, mental health, ageing populations, and indigenous communities like the Orang Asli. Socioeconomic issues, which are central to Malaysia's development, were found to be limited in corpus-based research. Malaysia has a unique sociocultural and political context. Greater scholarly attention towards issues such as poverty, inequality, social exclusion, and vulnerable communities warrants serious consideration as they continue to be addressed in national development agendas and public policy initiatives. Thus, future studies should delve into the

discursive portrayal of socioeconomic issues in Malaysian media and explore how such matters are represented to the public. Understanding the localised portrayals of these globally relevant issues can reveal how Malaysian media construct national narratives, shape public understanding, and reflect broader societal concerns.

In general, the findings of this review may guide future researchers in operationalising corpus-based approaches in studying Malaysian media discourse. Underrepresented themes, diverse corpus designs, and the integration of advanced corpus tools could be further examined and explored in future studies. Such efforts may contribute significantly to a more extensive, methodologically robust, and meaningful investigation of Malaysian media discourse.

Nevertheless, this review is subject to several limitations. Firstly, due to time constraints, this study was limited to only two databases, namely Scopus and Web of Science (WoS). Gusenbauer and Haddaway (2020) suggest various databases to ensure comprehensive coverage in systematic reviews. While Scopus and WoS are widely recognised for indexing high-impact, peer-reviewed journals, the exclusion of other databases may have led to the omission of relevant studies. Secondly, the scope of this review was narrowed to studies published between 2019 and 2024. Scoping to reviewed studies from recent years may have excluded earlier studies, which could bring analytical and theoretical depth to the analysis. Thirdly, this review focused only on English articles.

As articles in the Malay language were excluded from selection, this could skew crucial thematic findings, especially in socioeconomic themes in Malaysia. Lastly, since news reports play an influential role in shaping opinions among the audience, the reviewed studies focused on newspaper discourse. Consequently, other media forms, such as social media platforms and television broadcasts, were not reviewed in this study. Thus, the findings cannot be generalised to a wider Malaysian media discourse landscape.

ACKNOWLEDGEMENT

I am sincerely grateful to all who have supported this study, especially my supervisors, for their invaluable guidance and constant encouragement in completing this study.

REFERENCES

- Abas, N., & Kaur, S. (2019). Theologising the Discourse of Representation through 'Faith' and Religion in Football News. *3L: The Southeast Asian Journal of English Language Studies*, 25 (3), 150-162. ISSN 0128-5157. <https://doi.org/10.17576/3L-2019-2503-12>
- Abas, N., Abdul Aziz, R., Turiman, S., & Mat Daud, N.S. (2023). A corpus-assisted critical discourse analysis of COVID-19 vaccination-related news discourse in the Malaysian mainstream media. *3L: Language, Linguistics, Literature® The Southeast Asian Journal of English Language Studies*, 29(3), 148-165. <http://doi.org/10.17576/3L-2023-2903-11>
- Abdul Latip, A.R., & Othman, N. (2021). A socio-economic transformation of the ECER-development programmes in enhancing the well-being of the people. *GEOGRAFIA Online™*
- Malaysian Journal of Society and Space*, 17(3), 167-181. <https://doi.org/10.17576/geo-2021-1703-12>
- Abdul Malik, N., Ya Shak, M.S., & Hasni, N.A. (2022). A corpus-based approach to frame 'Covid-19 Vaccination' in Malaysian English newspapers. *Jurnal Komunikasi: Malaysian Journal of Communication*, 38(4), 23-42. <https://doi.org/10.17576/JKMJC-2022-3804-02>
- Abdul Malik, N., Ya Shak, M.S., & Hasni, N.A. (2023). Examining the framing of 'Covid-19 Vaccines': A corpus-based investigation of Malaysian newspapers. *Studies in English Language and Education*, 10(2), 1022-1040. <https://doi.org/10.24815/siele.v10i2.25883>
- Ahmed, S., & Matthes, J. (2017). Media representation of Muslims and Islam from 2000 to 2015: A meta-analysis. *International Communication Gazette*, 79(3), 219-244. <https://doi.org/10.1177/1748048516656305>
- Ariffin, A. S., Ismail, S., & Ali, S. S. (2022). Enhancing socioeconomic development spurred through STI policy framework into the nucleus off mainstream society in Malaysia. *Frontiers in Political Science*, 4, Article 878847. <https://doi.org/10.3389/fpos.2022.878847>
- Asad, S., Mohd Noor, S. N. F., & Jaes, L. (2019). Social actors' representation in online Malaysian newspapers during elections: A transitivity analysis of newspaper text. *Humanities & Social Sciences Reviews*, 7(4), 580-589. <https://doi.org/10.18510/hssr.2019.7478>
- Asad, S. (2022). The genre contrastive analysis of the first 14 weeks of Dr. Mahathir Mohamad in government: a study on two online newspapers under attitude analysis. *Revista de Investigaciones Universidad del Quindío*, 34(1), 292-311. <https://doi.org/10.33975/riuuq.vol34n1.798>
- Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a Curated, High-Quality Bibliometric Data Source for Academic Research in Quantitative Science Studies. *Quantitative*

- Science Studies*, 1, 377-386. https://doi.org/10.1162/qss_a_00019
- Bailey, C., Venta, A., & Langley, H. (2020). The bilingual [dis]advantage. *Language and Cognition* (12)2, 225-281. <https://doi.org/10.1017/langcog.2019.43>
- Baker, P., & McEnery, T. (2005). A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts. *J. Lang. Polit.* 4, 197-226. <https://doi.org/10.1075/jlp.4.2.04bak>
- Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.
- Baker, P. (2015). Does Britain need any more foreign doctors? Inter-analyst consistency and corpus-assisted (critical discourse analysis). In: Groom N, Charles M, John S (Eds.) *Corpora, grammar and discourse: in honour of Susan Hunston*. John Benjamins, Philadelphia, pp. 283-300.
- Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, 1(1), 363-376. https://doi.org/10.1162/qss_a_00018
- Brandtler, J. (2020). Vi bara testade en hypotes ::: Ännu mer om preverbal adverbial i svenska [We just tested a hypothesis ::: More about preverbal adverbials in Swedish]. *Norsk Lingvistisk Tidsskrift*, 38. 59-92.
- Copeland, D. A. (2003). *Newspapers in the Americas*. Praeger Publishers.
- Cotter, C. (2001). Discourse and Media. In D. S. al, *The Handbook of Discourse Analysis* (pp. 416-436). Blackwell Publishers.
- Cushion, S., & Thomas, R. (2013). The mediatization of politics: Interpreting the value of live vs. edited journalistic interventions in UK television news bulletins. *The International Journal of Press/Politics*, 18(3), 360-380. <https://doi.org/10.1177/1940161213484522>
- Chang, L. X., & Abdul Rahim, H. (2021). From positive to death: A corpus-based semantic analysis of COVID-19 representation in Malaysian English news reports. *3L: Language, Linguistics and Literature, The Southeast Asian Journal of English Language Studies*, 27(4), 16-28. <http://doi.org/10.17576/3L-2021-2704-02>
- Chouliaraki, L., & Fairclough, N. (1999). *Discourse in late modernity: Rethinking critical discourse analysis*. Edinburgh University Press. <https://doi.org/10.1515/9780748610839>
- Chukwunalu, Y. P., Nwankwere, U. N., Dereck, A., & Orji, M. S. (2022). A study of language use impact in radio broadcasting: A linguistic and big data integration approach. *Journal of Sensors*, 2022, Article 1440935. <https://doi.org/10.1155/2022/1440935>
- Entman, R. M. (1993). Framing: toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51-58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- Fairclough, N. (1995). *Media discourse*. Edward Arnold.
- Fairclough, N. (2013). *Critical discourse analysis. The critical study of language*. Routledge.
- Farrús, M. (2023). Automatic Speech Recognition in L2 Learning: A Review Based on PRISMA Methodology. *Languages* 8, 242. <https://doi.org/10.3390/languages8040242>
- Gomez, J., Mustafa, K. Anuar., & Yuen, B. L. (2018). *Media and elections: Democratic Transition in Malaysia*. Strategic Information and Research Development Centre.
- Gomide, A. R. (2020). Corpus linguistics software: understanding their usages and delivering two new tools. [Doctoral dissertation, Lancaster University] <https://core.ac.uk/reader/355104581>
- Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating

- retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research synthesis methods*, 11(2), 181-217. <https://doi.org/10.1002/jrsm.1378>
- Hajimia, H., Sarjit Singh, M. Kaur., & Rathakrishnan, M. (2019). Analyzing frequency and distributional pattern of discourse markers in the Malaysian newspaper. *Opcion*, 34(19), 1692-1714.
- Hardie, A. (2012). CQPweb—combining power, flexibility and usability in a corpus analysis tool. *International journal of corpus linguistics*, 17(3), 380-409. <https://doi.org/10.1075/ijcl.17.3.04har>
- Hua, Y. (2022). A critical overview of the major discourse-analytical approaches for media discourse studies. *International Journal of Research in Education Humanities and Commerce*, 3(5), 100-109.
- Ismail, H., Baharun, H., Abdullah, H., & Abd Majid, S.N. (2020a). Power and voices of authority in the media narrative of Malaysian natives: Combining corpus linguistics and discourse analysis approaches. *3L: The Southeast Asian Journal of English Language Studies*, 26(2), 50-66. <http://doi.org/10.17576/3L-2020-2602-04>
- Ismail, H., Idrus, M.M., & Syed Sahuri, S.N. (2020b). Masculinity through 'emotion': A corpus-assisted discourse analysis on sport news and implications for social workers in sport. *Journal of Human Behaviour in the Social Environment*, 30(1): 54-70. <https://doi.org/10.1080/10911359.2019.1664358>
- Jabeen, M., & Rashid, R. (2025). *Advancements in Corpus Analysis Tools: A Comprehensive Guide for Linguistic Research and Novice Users*. <https://doi.org/10.20944/preprints202502.0639.v1>
- Jeffries, L. (2010). *Critical stylistics: The power of English*. Palgrave Macmillan.
- Joharry, S.A., Turiman, A., & Mohd Noor, M. (2023). Comparative Collocation(al) Analysis between Two Malaysian Online Newspapers: A Case Study of the 2020 Sabah Election Day. *3L: Language, Linguistics, Literature@ The Southeast Asian Journal of English Language Studies*, 29(2), 146-168. <http://doi.org/10.17576/3L-2023-2902-11>
- Kilgariff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. & Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1(1), 7-36. <https://doi.org/10.1007/s40607-014-0009-9>
- Khalid, S., Kaur, S., & Lee, C. (2020). Media representation of online maid hiring system (SMO): A critical discourse analysis. *Media Education (Mediaobrazovanie)*, 60(3), 394-403. <https://doi.org/10.13187/me.2020.3.394>
- Kress, G. (2012). *Multimodality Analyses of Discourse in the Routledge Handbook of Analyses of Discourse*. (Eds.) James Paul Gee & Michael Handford. Routledge.
- Kwan, Y.K., & Ying, H.K. (2023). Seeking Legitimation in Political Uncertainties: Reforming the Media. In Rajandran, K. & Lee, C. (Eds.) *Discursive Approaches to Politics in Malaysia. Asia in Transition, Vol. 18*. (pp. 97-115). Springer. https://doi.org/10.1007/978-981-19-5334-7_6
- Leech, G. N. (2002). Corpora. In K. Malmkjær (Ed.), *The linguistics encyclopedia* (2nd ed., pp. 84-93). Routledge.
- Lockwood, C., Munn, Z., & Porritt, K. (2015). Qualitative research synthesis: methodological guidance for systematic reviewers utilising meta-aggregation. *International journal of evidence-based healthcare*, 13(3), 179-187. <https://doi.org/10.1097/XEB.0000000000000062>
- Masullo, C., Dentella, V., & Leivada, V. (2023). 73% of the observed bilingual (dis)advantageous effects on cognition stem from sociolinguistic factors: a systematic review. *Bilingualism: Language & Cognition*, First View, 1-15. <https://doi.org/10.1017/S1366728923000664>

- Mat Yakim, M., Mohamed, S., & Abdul Manan, K. (2019). Agenda Setting by The Malay Online Newspapers During the Semenyih By-election. *International Journal of Law, Government and Communication*, 4(17), 121-131. <https://doi.org/10.35631/ijlgc.4170012>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Mohd Ali, S.M., Wan Fakhruddin, W.F.W., Wan Adnan, W. N. A., & Azizan, E.A. (2024). Discursive construction of news reports on people with mental illness in Malaysian online newspapers. *Journal of Nusantara Studies*, 9(2), 371-401. <https://doi.org/10.24200/jonus.vol9iss2pp371-401>
- Mohd Nor, N. F., & Zulcaffi, 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>
- Mohd Nor, N.F., Jeffree, N. B., & Abdullah@Mohd Nor, H. (2021). Health is wealth: A corpus-driven analysis of the portrayal of mental health in Malaysian English online newspapers. *GEMA Online® Journal of Language Studies*, 21(2), 46-71. <http://doi.org/10.17576/gema-2021-2102-03>
- Murugaiah, N., Jamal, M., & Mehar Singh, M. K. (2024). Ideological Representations of People with Mental Illness in Malaysian Online Newspapers: A Critical Discourse Analysis. *GEMA Online® Journal of Language Studies*, 24(3), 53-72. <http://doi.org/10.17576/gema-2024-2403-04>
- McEnery, T. (2006). *Swearing in English: Bad Language, Purity and Power from 1586 to the Present*. London: Routledge.
- Nesi, H. (2012). ESP and corpus studies. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 407-426). John Wiley & Sons. <https://doi.org/10.1002/9781118339855.ch21>
- O'Halloran, K. (2010). How to use corpus linguistics in the study of media discourse. In A. O'Keeffe & McCarthy (Eds.), *The Routledge handbooks of corpus Linguistics* (pp. 563-577). Routledge.
- O'Keeffe, A., McCarthy, M. J., & Carter, R. A. (2007). From corpus to classroom. *Cambridge University Press*. <https://doi.org/10.1017/CBO9780511497650>
- Ong, C.T., & Hajar Abdul Rahim. (2019). Apprehending the culprits: The discourse of anti-profiteering and law enforcement in goods and services tax (GST) news reports. *Kajian Malaysia*, 37(2), 21-47. <https://doi.org/10.21315/km2019.37.2.2>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., ... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ (Clinical research ed.)*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Papaioannou, D., Sutton, A., Carroll, C., Booth, A., & Wong, R. (2010). Literature searching for social science systematic reviews: consideration of a range of search techniques. *Health Information and Libraries Journal*, 27(2), 114-122. <https://doi.org/10.1111/j.1471-1842.2009.00863.x>
- Preston, P. (2016). Inequality and liberal democracy: A critical take on economic and political power aspects. *Javnost—The Public*, 23(1), 37-54.
- Reisigl, M., & Wodak, R. (2001). *Discourse and discrimination. Rhetorics of racism and anti-Semitism*. Routledge.

- Roslani, M. A., & Jamal, M. (2024). (Mis) representation of the LGBT community through nomination and categorisation: The case of Nur Sajat in The Star. *3L: Language, Linguistics, Literature® The Southeast Asian Journal of English Language Studies*, 30(2), 63-77. <http://doi.org/10.17576/3L-2024-3002-05>
- Sadjirin, R., Abdul Aziz, R., Nordin, N.M., Ismail, M. R., & Baharum, N.D. (2018). The development of Malaysian corpus of financial English (MaCFE). *GEMA Online Journal of Language Studies*, 18(3), 73-100. <http://doi.org/10.17576/gema-2018-1803-05>
- Samaie, M., & Malmir, B. (2017). US news media portrayal of Islam and Muslims: A corpus-assisted critical discourse analysis. *Educ Philos Theory*, 49(14), 1351-1366. <https://doi.org/10.1080/00131857.2017.1281789>
- Selvaraj, S., & Sandaran, S. C. (2024). Discourses of Aging in a Malaysian Online Newspaper: Shaping the Perceptions of Society. *Research on Ageing and Social Policy*, 12(2), 105-126. <https://doi.org/10.17583/rasp.14369>
- Taufek, T. E., Mohd Nor. N. F., Jaludin, A., Tiun, S., Lam, K. C., & Aziz, J. (2022). Media Report about climate change in an English online Malaysian newspaper through thematic and discourse analysis approaches. *3L: Language, Linguistics, Literature*, 28(2), 193-210. <http://doi.org/10.17576/3L-2022-2802-1>
- Tiffen, R., Jones, P. K., Rowe, D., Aalberg, T., Coen, S., Curran, J., Hayashi, K., Iyengar, S., Mazzoleni, G., Papathanassopoulos, S., Rojas, H., & Soroka, S. (2014). Sources in the news: A comparative study. *Journalism Studies*, 15(4), 374-391. <https://doi.org/10.1080/1461670X.2013.831239>
- Valentine, K. D., Kopcha, T. J., & Vagle, M. D. (2018). Phenomenological Methodologies in the Field of Educational Communications and Technology. *TechTrends*, 62(5), 462-472. <https://doi.org/10.1080/10911359.2019.1664358>
- Van Dijk, T. A. (1991). *Racism and the Press*. Routledge.
- Van Leeuwen, T. (2008). *Discourse and Practice: New Tools for Critical Discourse Analysis*. Oxford University Press.
- Wang, D., & Liang, S. (2024). Discourse construction of Chinese modernisation from the perspective of Malaysian media. *Humanities and Social Sciences Communications 11*, Article 653. <https://doi.org/10.1057/s41599-024-03152-9>
- Wimmer, R. D., & Dominick, J. R. (2012). *Mass media research* (7th. ed.). Cengage Learning.
- Wu, H., A. Bakar, K., Jaludin, A., & Mat Awal, N. (2022). Sentiment analysis of China-related news in the Star Online Newspaper. *GEMA Online® Journal of Language Studies*, 22(3), 155-175. <http://doi.org/10.17576/gema-2022-2203-09>
- Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research*, 39(1), 93-112. <https://doi.org/10.1177/0739456X17723971>
- Xin, B., & Gao, X. L. (2019). *A Comparative Study of the Textual and Pragmatic Functions of Reported Speech in Chinese and American Newspapers*. Shanghai Foreign Language Education Press.
- Yoong, M., & Lee, S. (2023). 'China doll snatched away my husband': The intersectional othering of Chinese migrant women in a Malaysian newspaper. *Discourse, Context & Media* 54, 1-10. <https://doi.org/10.1016/j.dcm.2023.100713>
- Zhao, X. J., & Zhang, B. (2019). A study on the reported speech and construction of news discourse dynamic analysis framework. *For. Lang. Res.* 5, 37-42. <https://doi.org/10.16263/j.cnki.23-1071/h.2019.05.007>